X-RAY ANALYSIS AND PROTEIN STRUCTURE

By F. H. C. CRICK and J. C. KENDREW

The Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge, England

CONTENTS

PAGES

I. Introduction 134
II. The Nature of X-Ray Diffraction 135
1. Diffraction from a Crystal 135
a. Crystal Lattice and Reciprocal Lattice 137
b. Structure Determination 139
c. Symmetry 140
2. Diffraction from a Fiber 142
a. The Interpretation of Fiber Diagrams 144
3. Stereochemistry and Molecular Packing 146
4. General Remarks 151
5. Summary 153
III. Fibrous Proteins and Synthetic Polypeptides 154
1. Synthetic Polypeptides (and Silk) 154
a. α-Polypeptides 155
b. β-Polypeptides and Silk 158
c. Polyproline 159
d. Polyglycine 159
2. Fibrous Proteins 163
a. The α-Keratin Pattern 163
b. The β-Keratin Pattern 165
c. Collagen 166
IV. Crystalline Proteins 170
1. The Nature of Protein Crystals 170
2. Direct Information 174
a. Unit Cell and Space Group 174
b. Molecular Weight 175
c. Identification and Identity 176
d. The Shape of Protein Molecules 177
3. The Patterson Function 178
4. Methods Involving Heavy Atoms 179
a. The Heavy Atom Method 179
b. The Method of Isomorphous Replacement 180
c. Isomorphous Replacement and the Structure of Hemoglobin 183
d. Isomorphous Replacement and the Structure of Myoglobin 188

e. Isomorphous Replacement and the Structure of Ribonuclease 192
f. Requirements for Isomorphous Replacement 193
5. The Chain Configuration of Globular Proteins 197
V. Viruses 199
1. Rod-shaped Viruses 200
a. General Features of TMV 200
b. X-Ray Results: Basic Features 201
c. X-Ray Results: the Internal Structure 202
d. Correlation between X-Ray and Other Results 205
2. Spherical Viruses 207
a. Tomato Bushy Stunt Virus 207
b. Turnip Yellow Virus 207
3. General Principles of Virus Structure 208
References 210

I. Introduction

The last account of X-ray studies of proteins in *Advances in Protein Chemistry* was written by Fankuchen in 1945 (apart from Corey's article on Amino Acids and Peptides, in 1948). Dramatic progress has been made since then, particularly in studies of synthetic polypeptides and fibrous proteins; and although the goal of the protein crystallographer—the complete determination of the structure of a globular protein—remains unachieved, there have been considerable advances in method which should shortly pay dividends.

In this review we shall not try to cover the whole field but only the more successful and the more pregnant parts of it. One major omission is a description of the large-scale structure found in fibrous proteins such as collagen, the muscle proteins, etc. We have not discussed in detail results from other techniques such as infrared, optical rotation, etc., although we refer briefly to the results of such studies. The comprehensive articles by Low (1953) and by Kendrew (1954b) in *The Proteins* should be consulted for background material and for historical details, and for a summary of the most recent advances, a recent review by Kendrew and Perutz (1957).

We shall assume that the reader is familiar with proteins, but unfamiliar with crystallography. We shall not, therefore, set out crystallographic arguments in detail, but shall try rather to present a bird's eye view of the subject. This will enable the biochemist at least to catch the drift of crystallographic discussions; and also to gain some impression of which parts of the subject are speculative and which parts certain. Besides, protein crystallographers need help from biochemists; it will be part of our purpose to indicate where help is most needed.

We have called our review "X-Ray Analysis and Protein Structure." In last year's *Advances in Protein Chemistry* an excellent article appeared

by Anfinsen and Redfield (1956) entitled "Protein Structure in Relation to Function and Biosynthesis." Here we shall use the term "protein structure" in an entirely different sense—indeed, there is very little common ground between the two articles. Whereas Anfinsen and Redfield have concerned themselves with the amino acid sequence and topological interconnections of the polypeptide chains, we shall consider mainly the geometrical aspects—the arrangement of the atoms in space.

As the complexity of molecules increases, the geometrical aspects of structure become more and more important, and it is less and less possible to explain chemical behavior without taking into account dimensions and exact geometrical relationships. The strength of the X-ray approach to protein structure is that it alone among all the techniques available can hope to provide this precise quantitative information about molecules as complicated and as delicate as proteins. However, it is almost always an advantage in obtaining the geometrical structure of a molecule to know as much as possible about its chemistry; in particular a knowledge of the amino acid sequence of a protein should be a very considerable, if not indispensable, help to the crystallographer. It is likely to be a very long time before X-ray analysis can obtain by itself the amino acid sequence of a protein. Both methods—amino acid sequence determination and X-ray diffraction—will be necessary to obtain the complete structure, chemical and geometrical, of a protein.

II. The Nature of X-Ray Diffraction

X-ray diffraction is not a difficult branch of physics: on the contrary, it is easy to the point of tediousness. The widespread view that it is unintelligible has arisen because a certain intellectual effort is needed to grasp its mathematical foundations, and because it is supposed, incorrectly, that some special type of "three-dimensional imagination" is a prerequisite for understanding its methods and results. In this section we shall not attempt to expound the basic theory, but rather to characterize some of the broad features of X-ray diffraction, so that the reader may become familiar with the important concepts and with some of the jargon. To those who wish to build on a firmer foundation we recommend the more orthodox approach in such accounts as those of Bragg (1939), James (1950), Bunn (1945), and Robertson (1953).

1. Diffraction from a Crystal

All matter diffracts X-rays, but it is simplest for our purpose to start with diffraction from crystals. We mount a small crystal in a known orientation in the path of a fine beam of monochromatic X-rays; the X-rays scattered from it are caught on a photographic plate mounted some

distance behind the crystal. What does the picture on the plate look like? A rather nice example is given in Fig. 1. (To get this picture to look so pretty there had to be a good deal of "hokey-pokey"—complicated movements of the crystal and of the photographic plate, as well as a special screen.)

The reader will at once be struck by two features of the picture: its regularity and its symmetry. He should not be surprised by them, however, because regularity and symmetry are among the most important properties of crystals themselves. A crystal consists of a regular three-dimensional lattice of "unit cells," such that every unit cell has the same relation to its neighbors; and the contents of every unit cell are the same, namely a number of identical molecules related to one another by symmetry elements, such as rotation or screw axes or planes of symmetry.

In an analogous manner, the X-ray picture exhibits regularity, which lies in the positions of the spots. They form a regular two-dimensional array or lattice. That is to say, the distance between a spot and its neighbors is the same no matter where on the picture the spot may be. The only exceptions are those cases where the spot is too weak to be shown on the photograph; and this draws our attention to two other features of the picture. First, while the positions of the spots are regular, their blacknesses (their intensities, that is) differ—some are strong, some weak. Second, the pattern exhibits symmetry; that is to say, it can be divided into four quarters which are identical. Other X-ray photographs of this sort might have shown different types of symmetry, but some there would always be.

What, then, is the significance of each particular spot at such and such a position on the photograph and of such and such a blackness? What, in fact, is it that scatters the X-rays? To the last question we can give a simple answer: it is electrons which scatter X-rays, in this case the electrons of the atoms in the crystal. Atoms have finite sizes and electrons distribute themselves over atoms and the bonds which join them. It is convenient to think of a crystal as a three-dimensional pattern of *electron density* which reaches high values near the centers of atoms and low or zero values in the spaces between. It can be shown that the X-ray picture represents a "wave analysis" (sometimes called Fourier analysis) of this electron density. When we make a wave analysis of a crystal we think of it as made up of a very large number of waves of electron density, running[1] in many different directions through it. If we have carried out the analysis correctly, we shall find that when we add together all these waves—each of the correct size (amplitude) and to the right extent in or out of step with its neighbors (phase)—we get back to the actual electron

[1] But, like the Red Queen, they run without getting anywhere.

density of the crystal. This is a three-dimensional wave analysis, often known as a Fourier analysis, and the reverse process is a wave (or Fourier) synthesis. We are familiar with analogous processes involving only one dimension. In music, a harmonic analysis of the profile of sound produced by, say, a violin playing a steady note, gives us a fundamental and a series of harmonics, each separately being a simple or sinusoidal wave, and all of them adding up (synthesizing) to re-form the original profile.

The significance of a particular X-ray spot is that it corresponds to *one* of these (imaginary) sinusoidal waves of electron density. The *position* of the spot on the picture shows us both the direction of the wave and its wavelength. If the spot is near the center of the picture, the corresponding wave of electron density is one having a large wavelength. If it is far from the center it corresponds to a wave of short wavelength. Thus the *outer parts* of X-ray pictures are concerned with fine details of structure (high resolution), the *inner parts* with broad features (low resolution). The direction of the spot, relative to the center of the picture, shows the direction of the electron density wave. Thus a spot *vertically* above the center corresponds to a wave of electron density in the crystal whose direction is *vertical*—i.e., to horizontal layers of high electron density separated by regions of low electron density. The intensity of the spot is related to the amplitude of the wave—in fact the square of the amplitude is proportional to the intensity of blackening of the plate—and an intense spot implies that this particular electron density wave must be of large amplitude for the structure in question.

a. Crystal Lattice and Reciprocal Lattice. In this section we shall explain the concept of the "reciprocal lattice," which is nothing more than an abstract way of representing the diffraction pattern. A crystal is a three-dimensional structure, but the picture in Fig. 1 is clearly a two-dimensional affair; and we must now confess that it contains not the whole diffraction pattern of a crystal but only part of it. The reader is to imagine that the complete pattern is a *three*-dimensional lattice of X-ray spots, of which Fig. 1 is just one particular plane—actually a plane going through the origin. This three-dimensional lattice is known as the *reciprocal lattice* of the crystal, and it is important to have a general picture of its properties. X-ray cameras are merely devices which allow a part of the three-dimensional reciprocal lattice to be recorded on a two-dimensional photographic plate in a systematic way so that spots can readily be identified ("indexed"). Different types of cameras may therefore present the same array of reflections arranged in different ways; but whichever way one takes the picture there is one characteristic of an X-ray spot which can always be obtained directly from it. This is its "spacing"; that is to say, the wavelength of the imaginary wave of electron density ("Fourier component") to which it corresponds.

As we have already pointed out, the *larger* the distance of the spot from the center of the picture (that is from the origin of the reciprocal lattice), the *smaller* the "spacing"—hence the word reciprocal. And the stronger a given spot, the larger the amplitude of the corresponding Fourier component; in other words, the larger the number of electrons (and therefore of atoms) clustered near the anti-nodes of the wave. For example, a

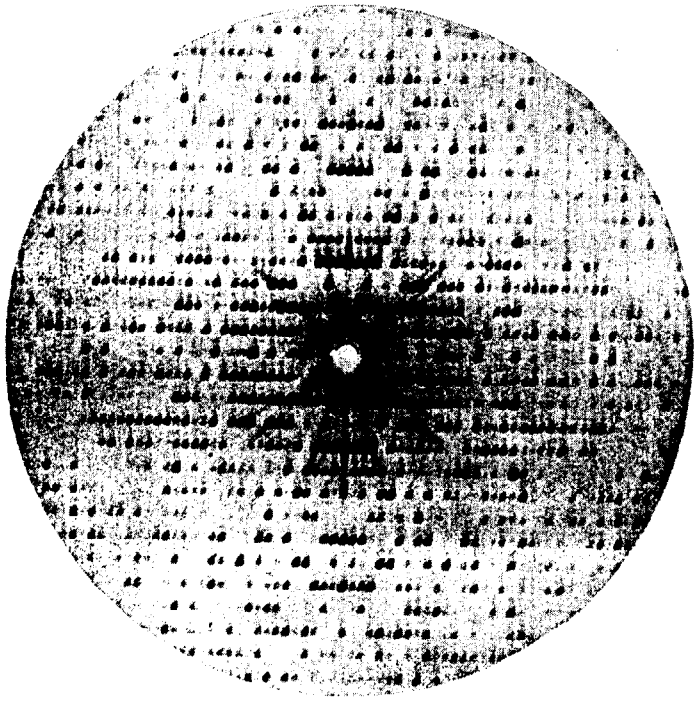

FIG. 1. A typical X-ray photograph of a protein crystal. Notice that the spots form a regular two-dimensional lattice; note also the symmetry. The picture shows only a small part of the complete X-ray diffraction pattern of the crystal. (After Kendrew and Kraut: finback whale myoglobin, type F, *c* projection).

strong spot above the origin (*meridional*) with a spacing of 1.5 A. implies that in the crystal there are certain horizontal parallel planes 1.5 A. apart, near which many atoms cluster. Finally, from the regular dimensions of the reciprocal lattice one can directly calculate the dimensions of the repeating unit, or unit cell of the crystal—and because of the reciprocal relationship, the smaller the unit cell the greater the distance apart of the spots in the reciprocal lattice. To recapitulate our musical analogy, the unit cell dimensions are the three-dimensional analog of the wavelength of the "fundamental tone" in a musical sound.

b. Structure Determination. The relationships between the *real* lattice—that is, the real three-dimensional crystal or, rather, its three-dimensional repeating pattern of electron density—and the reciprocal lattice (or X-ray pattern) are very intimate ones. Given the position of all the atoms in the unit cell of a crystal it is a straightforward, if sometimes lengthy, matter to calculate the entire diffraction pattern of the crystal. This is interesting, but not often useful. It is the reverse process, given the diffraction pattern to discover the structure, which one more often has to contend with. Unfortunately it is by no means so simple to carry out.

There is, in fact, a fundamental reason why one cannot calculate the unknown structure from the experimental data merely by the use of some mathematical sausage machine such as a high-speed computer. In order to combine correctly all the (imaginary) waves of electron density which build up the correct structure it is necessary to know not only the *amplitude* of each of them (which one obtains from the blackness of the corresponding spot in the picture) but also its *phase*—that is to say, how far each train of waves is out of step with its neighbors—and this information is *not* given by the experimental data. In other words, *the experimental data contain just half the required information.* One has, therefore, the curious situation that if the structure can be correctly guessed one can check it against the X-ray data in a straightforward way; but the structure cannot be deduced from the data in a routine manner except in certain special, and very simple cases. There are, however, various stratagems which allow one sometimes to make a rather good guess at the structure—especially if one knows something about it before one starts; but guesswork is always involved, and it is this which makes crystallography something of an art. The pursuit of a structure is rather like hunting: it requires some skill, a knowledge of the victim's habits, and a certain amount of low cunning.

A number of the stratagems useful for solving structures will be mentioned later. Often the most useful is sheer intuition, based on experience and on what the chemists have already discovered about the formula of the molecule. For structures of moderate complexity the most powerful is probably the Patterson synthesis, whose properties and applications in the biological field have been fully described by one of us (Kendrew and Perutz, 1949; Kendrew, 1954a). Briefly, it is a method which does not involve guesswork, of presenting all the X-ray data in such a form as to display the *relative*, but not the absolute, positions of pairs of atoms in the structure. This may enable one to obtain important clues about the structure.

Then there is the method of isomorphous replacement, and its close relation, the heavy-atom method. The former involves comparing a crystal with a heavy atom in it with a very similar crystal not containing the heavy atom. The latter gives a first approximation to the structure calculated on the assumption that the heavy atom, whose position in the structure must be known, effectively swamps the rest of the atoms and determines the phases of all the reflections. In both methods we may

say loosely that the heavy atom acts as a marker in the crystal. They will be described more fully later in the article.

Recently a number of claims have been made that phases can be directly and deductively obtained by certain complicated mathematical procedures. This, if really so, would take the guesswork out of crystallography and would make the italicized sentence above untrue. These methods certainly work for simple structures, but so far no structure has been solved with their help which could not have been solved by the older methods. Moreover, there are good reasons for believing that this approach will not work when there are many atoms in the unit cell, as there always are in protein crystals.

c. *Symmetry.* The reader will be familiar with the fact that most crystals have symmetry elements, such as rotation axes and mirror planes. The symmetry of protein crystals is simpler than that of crystals in general, because certain types of symmetry elements—those which involve reflection (mirror planes, glide planes, and centers of symmetry)—are forbidden to them. This is because proteins are made up of optically-active amino acids all of which have the *levo*-configuration. A *levo*-compound, acted upon by a mirror plane, a glide plane, or a center of symmetry, gives a *dextro*-compound (just as a left-hand glove gives a right-hand glove); since dextro amino acids are not present in the crystal, these symmetry elements cannot be present either. This leaves rotation axes and screw axes as the only permitted symmetry elements for protein crystals. These two terms are formally defined as follows:

a. A crystal has an n-fold *rotation* axis if the structure appears identically the same after being rotated through an angle of $360°/n$ about the axis.

b. A crystal has an n-fold *screw* axis if the structure appears identically the same after first rotating through an angle $360°/n$ about the axis, and then translating it a certain distance parallel to the axis.

It is well known that for a crystal the only axes possible are 2-, 3-, 4-, or 6-fold, whether they be rotation axes or screw axes.

There is a very close relation between the symmetry of the crystal lattice (or real lattice) and that of the reciprocal lattice, although the symmetry of the reciprocal lattice is always higher. In protein crystals, if the reciprocal lattice has an n-fold axis of symmetry, then the real lattice must have either n-fold rotation axes or n-fold screw axes (or both) in the same direction. (In addition the reciprocal lattice has a center of symmetry, which, as we have already said, cannot occur in the real lattice of a protein crystal). In the X-ray picture a screw axis can usually be distinguished from a rotation axis, since the former always causes certain X-ray spots to have zero intensity systematically, whereas the latter can only do this fortuitously. It is thus usually possible to determine unambiguously not only the size of the unit cell, but also all the symmetry elements which are present. This assembly of symmetry elements is called the *space group* of the crystal. Notice that the space group does not depend on the

size of the unit cell, but only on its symmetry elements. It is possible to show by quite general arguments that only 230 different space groups are possible for crystals, and of these only 69 need trouble the protein crystallographer—all the others involve forbidden symmetry elements. An analogous term—point group—is useful in discussing virus structure. It refers to the symmetry elements possessed by an arrangement which is *finite* in all directions, and therefore clusters around a point.

There is one more piece of jargon which we must introduce at this

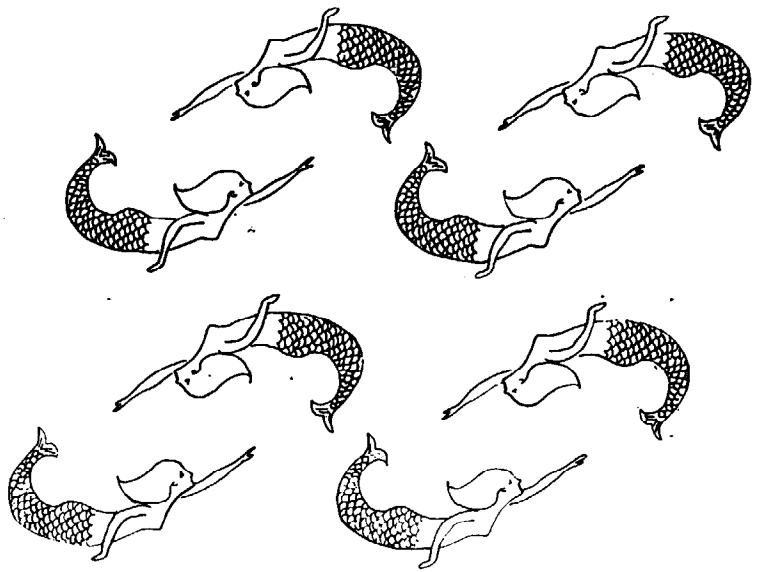

FIG. 2. An example from everyday life to illustrate the difference between unit cell and asymmetric unit. There are two mermaids in the unit cell, but only one in the asymmetric unit. Notice that this pattern is the same upside down.

stage: the *asymmetric unit.* This the the smallest part of the structure which, when operated upon by the symmetry elements of the space group, will reproduce the complete structure. It is thus a smaller unit than the unit cell, which may contain several asymmetric units, just as it contains a complete set of symmetry elements. Asymmetric units are related by the symmetry elements: several asymmetric units make up the unit cell, and unit cells are related by translations. The distinction will be easy to grasp from a diagram (Fig. 2). We may say that the asymmetric unit is the brick from which the structure is built up, but it is a crystallographer's brick and not necessarily the same as the chemist's brick, which is the molecule. The asymmetric unit is often the same as the chemist's molecule, but it may sometimes be bigger and sometimes smaller. Thus the

asymmetric unit may contain two (occasionally more) molecules not related by symmetry elements of the space group (if they were so related, the asymmetric unit would be half the size and would contain one molecule); or, if the molecule consists of two identical subunits, each subunit may be a single asymmetric unit. The latter state of affairs exists in horse hemoglobin crystals. Notice that the environment of each asymmetric unit is the same as that of any other (see Fig. 2), so that if it contains one molecule it follows that every molecule in the crystal has an identical environment. If, however, the asymmetric unit contains more than one molecule, it follows that all the molecules do not have the same environments; and the X-rays do not recognize in a simple manner that these identical but differently arranged molecules really have the same structure. This will only show up when such a structure is completely solved.

2. Diffraction from a Fiber

The diffraction of X-rays by a fiber is not different in principle from diffraction by a crystal, but there are a certain number of differences in practice.

A fiber is not a single crystal. It is best thought of as a collection of small crystallites, generally embedded in a certain amount of amorphous material. Fibers are usually built up of polymers, that is, of long molecules constructed by the indefinite repetition of identical monomer units. Thus a single chemical molecule may run through several crystallites; the monomer in fibers corresponds to the molecule in single crystals. In a well oriented fiber the crystallites all lie with one axis (the "fiber axis") almost parallel to the length of the fiber, but the orientations around this direction are random or nearly so; if their orientations were all the same the structure would revert to a single crystal.

It follows that if we take an X-ray picture of a fiber the result will be similar to what we would get if we photographed a single crystal and continually rotated it about one axis during the exposure (Fig. 3). This means that instead of photographing one part of the reciprocal lattice at a time one obtains almost the whole of the three-dimensional reciprocal lattice on the same (two-dimensional) photographic plate. Thus it is not always easy to unscramble the X-ray picture and so to obtain an exact idea of the reciprocal lattice which produced it. Compare the difficulty of visualizing a person from a series of superimposed snapshots taken while he stood on a revolving table, in spite of which certain features could easily be established - for example one could see that the subject's eyes were above his mouth.[2]

[2] Unless, of course, he were standing on his head. Fortunately it does not matter if crystallographers stand on their heads; as we have already pointed out, the reciprocal lattice always has a center of symmetry so that X-ray pictures look just the same if studied from this posture.

The information which can most easily be obtained from a good fiber photograph is the crystallographic repeat in the direction of the fiber axis—this is shown by the spacing of the meridional reflections[3]; in favorable cases it may be possible to deduce the other dimensions of the unit cell as well. The symmetry is often difficult to deduce directly, but can sometimes be inferred from the dimensions of the unit cell. Finally, it is usually possible to get some idea about where in reciprocal space the strong reflections occur, and if the unit cell has been identified, to locate them precisely in the reciprocal lattice.

In a poor fiber the crystallites are only approximately parallel to the fiber axis, and this will cause the X-ray spots to be drawn out into circular

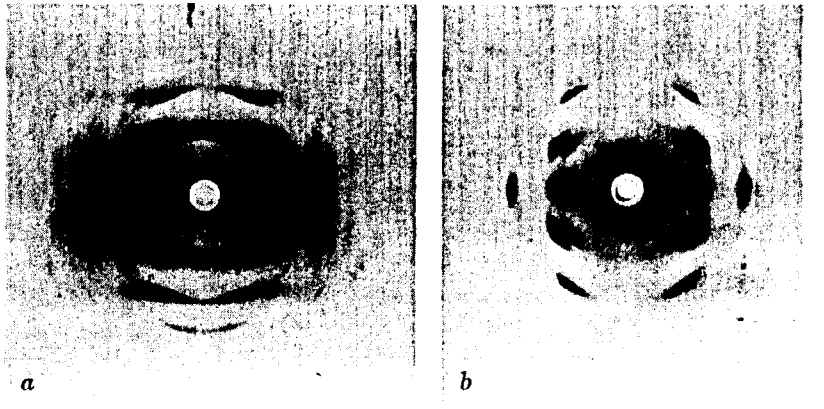

FIG. 3. (a) Poly-L-alanine, α form, (b) poly-L-alanine, β form. Two typical X-ray fiber diagrams, of good quality, showing many discrete spots. (Brown and Trotter, 1956).

arcs which makes it more difficult to locate them in reciprocal space, although the spacing of the spot (given by the radius of the arc) can always be measured. In practice all fiber photographs show this effect which also makes it more difficult to get an accurate measure of the intensities of the reflections.

What we have so far been describing is the best type of fiber photograph. Very often they are less well behaved than the specimen illustrated in Fig. 3. For a start, fiber diagrams rarely extend so far in reciprocal

[3] Conventionally the fiber is always mounted vertically, so that *vertical* (meridional) reflections correspond to spacings along the fiber axis, and *horizontal* (equatorial) reflections to spacings perpendicular to the fiber axis.

space as good single crystal photographs, the X-ray intensities fading away in the outer parts of the picture. The photograph is often confused by diffuse background scatter of X-rays from amorphous parts of the fiber. Even worse trouble comes if the crystallites themselves are disordered. We have so far considered the case where small regions of fiber exist which, within themselves, are perfectly crystalline, in three dimensions, and these we have called crystallites; or, in other words, we have local three-dimensional order. A common disorder is that of chain direction—chains may run upward or downward at random within an otherwise regular lattice. But there may be no three-dimensional order at all—only one-dimensional order; that is to say, although the polymer molecules all run (approximately) parallel to the fiber axis, there is no

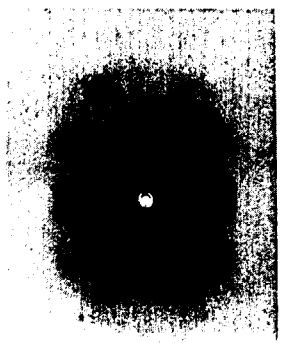

FIG. 4. The X-ray fiber diagram given by collagen (compare Fig. 3). It is a considerable technical achievement to get a picture of collagen to look even as good as this. (Dry rat tail tendon, stretched 8%. Cowan, North, and Randall, 1955).

correspondence between neighboring chains, which may be displaced up or down relative to one another in a random way. A fiber disordered in this way still shows the crystallographic repeat in the fiber direction, and the X-ray scattering produces layer-lines, or horizontal streaks, on the photograph. But there are no discrete *spots* on the layers (with one exception); instead there is a continuous variation of scattered intensity along each line. Surprisingly enough in some circumstances this may be an advantage rather than the reverse, providing even more information than a photograph consisting of discrete spots. An example of such a photograph is shown in Fig. 4.

a. The Interpretation of Fiber Diagrams. X-ray pictures of fibers are often too poor to allow one to use the methods of analysis customary for single crystals: in particular one usually cannot hope to "see" the atoms even when the correct structure has been discovered. The interpretation of fiber diagrams is a special art, therefore. The method of attack is to

try to deduce the symmetry of the fiber molecule from the X-ray picture; then to build scale models having this symmetry; and finally to show that only one of these models will fit all the available data, X-ray or other. Thus unless one knows in advance the chemical formula, or at least its most important features, the problem is almost hopeless. In addition, information derived by other techniques such as measurements of infrared dichroism is often invaluable.

The symmetry of a fiber molecule is almost always a screw axis. The reasons for this are explained in the next section, where it is pointed out that there is no reason why a single fiber molecule should not have a nonintegral screw axis. For example, 3.6 monomer residues per turn is 18 residues in 5 turns. Such a structure will have a "true repeat" after 18 residues, but this is not a very fundamental characteristic of it, since a very small twist of the molecule would give a different "true repeat" or even no true repeat at all. (Thus in our example a twist from 3.60 to 3.61 residues per turn would lead to a repeat of 65 residues in 18 turns.)

It might be thought that the absence of a short repeat distance would make the problem impossibly difficult to solve, but fortunately helical symmetry often produces striking effects in the photograph which immediately reveal its existence even when the screw axis is nonintegral. Until a few years ago the theory of the effects produced by nonintegral helices in crystal diffraction patterns had not been worked out, simply because in ordinary crystallography there is no occasion for it. It was in fact only developed (by Cochran *et al.*, 1952) in response to the proposal by Pauling and Corey (1951a) that the α-polypeptides were built up of nonintegral helices, namely the now famous α-helix with its 3.6 residues per turn and 1.5 A. per residue to which we have already several times implicitly referred. Armed with the appropriate theory it is often possible to recognize the helical nature of a fiber structure at a glance, and sometimes to specify the main parameters of the helix and its subunits with very little trouble indeed.

There is a catch, however. Imagine that a sheet of paper has been folded around the structure in the form of a cylinder, and a mark put on the paper at corresponding points in each asymmetric unit. If this paper is now opened out we shall obtain a pattern of the type shown in Fig. 5, which we shall call a *net-diagram*. Now what our helix theory is giving us is in essence the net-diagram of the structure, or at least one of a small number of possible net-diagrams. The positions of the points of a particular net are unambiguously determined, but not how the atoms inside each of them are arranged, nor, what is more important from the present point of view, how each net-point is *chemically* attached to its neighbors. Thus the net-diagram of Fig. 5 might correspond to any of the three arrangements shown by the arrows, or indeed to an infinite number of others.

Note that some of these possible arrangements have a single chain of subunits winding upward, some more than one chain.

To solve the structure completely, and thus resolve this ambiguity in the net-pattern, it is usually necessary to build models—the X-ray data alone are not sufficiently restrictive, and one's knowledge of chemistry must be invoked to fill the gap. Generally the chemical nature of the subunit is known (in polypeptides it is simply —NH—CO—CHR—), and, as indicated in the next section, many detailed stereochemical data are available from the literature. Armed with all this information, together

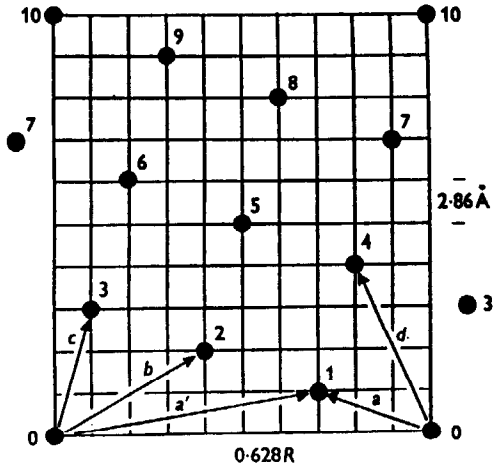

FIG. 5. The helix-net derived from wide-angle diagrams of collagen. Black dots represent the relative locations of "centers" of equivalent groups of atoms on a cylindrical shell of radius R (with axis vertical). The vectors a to d show several possibilities for connecting the black dots by means of polypeptide chains to form helical structures. The recent models of collagen all use connection c. As can be seen by studying the figure this connection corresponds to *three* separate chains winding round the same axis. (One chain joins 0, 3, 6, 9; another 2, 5, 8, and the third 1, 4, 7, 10.) (Bear, 1955.)

with any derived from subsidiary techniques, it is possible, with experience, to devise a scheme of systematic model building which will enable one to eliminate nearly all the infinite number of theoretical ways of joining up the points on the net-diagram. If all but one of these ways can be eliminated, and if its theoretical diffraction pattern gives reasonable agreement with the observed X-ray picture, the structure is essentially solved.

3. Stereochemistry and Molecular Packing

In this section we shall explain some of the conditions which *any* successful structure must satisfy. We shall also mention some of the principles underlying the construction of *regular* structures.

In any postulated structure the bond distances and bond angles must have acceptable values. The values to be regarded as acceptable are those derived from X-ray studies of small molecules, such as amino acids and small peptides, and the chance of any large deviations from the average values is negligible. In the field of proteins much of the work of deriving a canonical set of dimensions has been done at the California Institute of

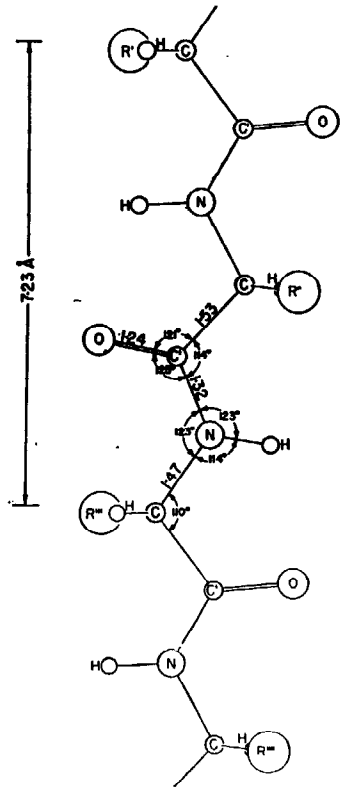

FIG. 6. A diagrammatic representation of a fully extended polypeptide chain with the bond lengths and bond angles derived from crystal structures and other experimental evidence. (Corey and Pauling, 1953.)

Technology, and the standard values for the polypeptide chain are those given by Corey and Pauling (1953) and shown in Fig. 6. Their most important feature is that the six atoms of the peptide group (—C—CO—NH—C—) invariably lie in a plane, or very nearly so. This is attributed to resonance, which is also responsible for the relatively short HN—CO bond. The structure retains freedom of motion in spite of the planarity of the peptide bond, rotation being possible about the two single bonds attached to each C_α atom.

Apart from the covalent bonds—usually known in advance from chemical studies—the most important links in structures of the type we shall be considering are the hydrogen bonds, such as NH. . .OC or OH. . .OC. Experience has shown that in practice virtually all the NH, CO, OH or similar groups which the structure contains are somehow linked up in hydrogen bonds; but the particular groups paired to one another cannot be predicted. The stereochemical conditions which a hydrogen bond must satisfy are not so restrictive as those governing covalent bonds, but the bond distance and bond angle must fall nevertheless within certain limits, which have been discussed by Donohue (1952). Finally there are the van der Waals' contacts between neighboring atoms. These are not directional bonds, nor are the permissible distances very precisely determined. However, a structure must not have van der Waals' contacts which are unacceptably short. All in all the conditions imposed by stereochemistry are very severe, and the number of configurations allowed by them for a structure is often very small. This does not necessarily mean that all the allowed configurations can be simply discovered.

It is considered nowadays good practice, when proposing a structure for a fibrous molecule, to give coordinates for its atoms to the nearest 0.1 A., or preferably 0.01 A. This does not imply that the author thinks he knows the coordinates as accurately as this; it merely indicates that a configuration giving acceptable bond distances and angles is at least possible. Specification of the exact coordinates allows this to be checked by others. The fact that the coordinates may be slightly wrong is not a valid excuse for failing to present a consistent set of them.

Apart from these stereochemical considerations the most important general principle in structural work is symmetry. It is a good working rule that where possible the same packing arrangements will be used over and over again in a structure. It follows that generally there will be symmetry elements of one sort or another, and since the presence of these can often be deduced rather directly from the X-ray data they are of considerable importance in tackling a structure. Of course, as we have already indicated, true crystals almost always possess symmetry elements. But this is often true of polymer molecules too, especially if they have been encouraged to take up a regular configuration by drawing them out into fibers. Otherwise they are called amorphous, and are then not very suitable for study by X-rays (see the next section).

If a fiber structure does repeat, the most likely symmetry element is a screw axis; a pure translation can be thought of as a special case of a screw axis with zero rotation, and is comparatively rare. Other symmetry elements (mirror and glide planes) are theoretically possible but are most improbable in practice; indeed they are impossible if the polymer contains asymmetric carbon atoms of only one hand. There are only two exceptions—if there are several chains in the structural unit they may be related by a rotation axis parallel to the axis of the fiber; and there is a possibility of dyad axes perpendicular to the fiber axis (as in deoxyribonucleic acid (DNA)). But usually such symmetry elements occur, if at all, in addition to a screw axis.

Unless it be a simple dyad, a screw axis generally gives a structure a helical appearance. Helices are ubiquitous in biology precisely because biological structures are very often made of small units linked together, end to end, to make up a larger entity. In recent years it has been realized that a single isolated helix may have a screw axis which is noncrystallographic; that is to say, n (as defined on p. 140) is not restricted to 2, 3, 4, or 6, but may assume any value, integral or nonintegral. A nonintegral value means merely that a single turn of the helix contains a nonintegral number of subunits. Note that in an *isolated* helix the environment of each subunit is the same whether n is integral or nonintegral; and there is no reason why a nonintegral value, such as 3.6, should not be assumed if packing relationships between neighboring residues in the helix are best satisfied in this way.

If, however, an attempt is made to pack such helices into a regular lattice, the relationship between asymmetric units in neighboring helices will not be the same everywhere unless the screw axis is 2, 3, 4 or 6-fold; in other words a true crystal cannot be formed unless this condition is satisfied, because it can be shown that these are the only symmetry axes (rotation or screw) which allow a pattern to repeat indefinitely in two or three dimensions. True, the chain molecule in a crystal may have a 3.6-fold axis of symmetry, but this symmetry cannot be apparent in the relations between it and its neighbors—it is accidental from the point of view of the crystal and cannot form part of the space group. Such a situation could only arise if the interactions between neighboring chains were relatively weak. *A nonintegral screw axis is likely to appear when the interactions of a fiber molecule with itself are much stronger than its interactions with its neighbors.* In crystallographic jargon a nonintegral screw axis is a pseudo-axis. It need not even apply to the whole of the fiber molecule. Thus the backbone of a polypeptide chain might have nonintegral screw symmetry, but not the distal ends of the side chains which are largely influenced by their neighbors.

If a small number of chains, each with a nonintegral screw axis, is placed side by side, they may try to interact in a regular manner, for example, by forming interchain hydrogen bonds. If they are to remain strictly parallel, regular interaction will not usually be possible since the nonintegral axis will cause the interchain bonds to get out of step. Sometimes, however, it may happen that if the individual (helical) chains coil slowly

around each other there is again a possibility of regular interlinking, the small additional twist bringing the chains into step. Such a distortion destroys the exactness of the helical configuration of the individual chains, but it is often so slight that the chemical bonds which brought that configuration about are not appreciably distorted. Such a structure is known as a superhelix or coiled coil, and an example of it is probably α-keratin (see Fig. 13). The smaller, primary, helix is referred to as the minor helix, and the gently helical path followed by its axis is called the major helix.

Nonintegral screw axes are not found only among fibrous molecules. An interesting example, recently discovered, is the rod-shaped tobacco mosaic virus (TMV). Here the asymmetric unit is not a single amino acid residue, but the whole of a globular protein molecule of molecular weight about 17,000. In other words the virus particle consists in the main of a large number of identical protein molecules stacked in a helical array. The asymmetric unit is a single one of these molecules, which thus all have the same environment.

We have so far considered symmetrical arrangements of subunits which are theoretically infinite in extent. In other words, as far as their symmetries are concerned an α-helix or a molecule of TMV might go on forever. In fact, of course, they do not do so. It is not clear what it is that causes them to terminate at a particular point, but it certainly is not the requirements of symmetry. We shall conclude by making brief reference to symmetrical arrangements of subunits which are *finite* in extent, since such arrangements have been shown very recently to be relevant to the structures of spherical viruses, as we shall indicate in Section V. The restrictions on the symmetry elements allowed in such arrangements are more stringent than ever; only rotation axes are permissible if the subunits are nonenantiomorphous (optically active). (It is not difficult to see that screw axes generate new asymmetric units *ad infinitum*—owing to the elements of translation involved—and must be inadmissible in a finite system.) Three general types of *point group*, or finite collection of symmetry elements, (see p. 141) are possible: first, those consisting only of an n-fold rotation axis (n = any integer); second, those possessing in addition dyad axes perpendicular to the main axis; and third, the cubic point groups. It is the latter which are important in the present connection because they generate isodimensional arrangements, such as spherical viruses are known to be. There are three cubic point groups which will interest us. The first has four threefold axes, arranged tetrahedrally, and a number of dyad axes. The second has fourfold, and the third fivefold axes as well as three- and twofold axes. The properties of these three

point groups, known as 23, 432, and 532 respectively,[4] are set out in Table I, together with the number of asymmetric units in each and the names of the regular (or Platonic) solids which possess these symmetry elements (among others).

TABLE I

The Three Non-Enantiomorphous Cubic Point-Groups

Crystallographic description	Number and type of rotation axes	Number of asymmetric units	Regular solids possessing the same symmetry elements
23	3 dyad 4 triad	12	Tetrahedron
432	6 dyad 4 triad 3 tetrad	24	Cube Octahedron
532	15 dyad 10 triad 6 pentad	60	Dodecahedron Icosahedron

4. General Remarks

In this section we shall briefly consider which aspects of a structure are most clearly "seen" by the X-rays. This should help the reader to cultivate his "X-ray eye," lack of which has so often caused misunderstanding in the past.

We have spoken so far as if X-rays are scattered only by repeating structures such as crystals; but this was a simplification, merely for didactic purposes. The fact is that X-rays are scattered by *every* part of the specimen, but *there will only be sharp spots on the X-ray photograph if the electron density is periodic in space.* Otherwise the photograph will show smears, smudges, or merely diffuse blackening. If part of the structure is periodic, part aperiodic, then there will be sharp spots superposed on smudges or diffuse background. Since a given amount of blackening shows up much more clearly if it is collected into a spot than if it is spread over an area, and since spots are easier to interpret, our attention is usually concentrated on them rather than on the smudges. So when we "solve a structure" we are generally describing those parts of it which are regular, that is, which repeat periodically in space.

Suppose we have a structure which does for the most part repeat regularly, with the exception that one small part of the unit cell is irregular,

[4] Pronounced two-three, four-three-two, and five-three-two.

varying in a random manner from cell to cell. What will the X-rays see? Strictly speaking, of course, they respond to the entire structure, but the X-ray *spots* "see" only the *average* unit cell. Such a situation may be met with in fibers, which may have a random arrangement of side chains; and also in crystalline proteins, where much of the solvent in the unit cell may have no regular structure. It is true in some measure of *all* X-ray photographs owing to the random thermal motions of the atoms in all crystals. All these effects to some extent smear out the average electron density and reduce the amount of fine detail which we can expect to see—which means that those parts of the reciprocal lattice corresponding to small spacings and fine details (the parts far from the center, that is to say) are reduced in intensity. In protein work the trouble is particularly acute, and the X-ray intensities from a protein crystal or fiber fall off with decreasing spacing much more rapidly than do those from an ordinary organic crystal; so the atoms in the structure of a protein, if we eventually succeed in "seeing" them, will certainly be smeared out somewhat. This will naturally make the interpretation of the results more difficult, even if they are known to be correct.

What would be the effect on the diffraction pattern of minor variations in the amino acid composition of the protein—a change in a single side chain, for example? Strictly speaking, almost every X-ray reflection is influenced to some extent by *every* electron in the unit cell. But a change in the few atoms making up a single side chain represents a very small change in the electron density distribution in the unit cell as a whole (protein side chains all have about the same electron density, and we may assume that they are generally packed close together without leaving any gaps which cannot at once be occupied by water molecules); hence the average effect of such a change on any given reflection is slight, generally well within the error of measurement. Small changes in a few reflections may, however, be just observable. It follows that we cannot expect to show by X-rays in any simple manner whether or not two very similar proteins are in fact identical.

X-rays see electron density, not atoms and bonds. Therefore they see a structure in terms of electron density and not as a chemist would see it. For example, they cannot even distinguish in a simple way where one molecule ends and the next begins—one must deduce this indirectly from a knowledge of bond dimensions. On the other hand they are very sensitive to even slight changes in the position or orientation of a molecule within the unit cell; changes of a kind to which protein crystals are peculiarly susceptible. Difficulties of this kind are not important when the correct three-dimensional electron density map of a structure has been obtained, but they have to be borne in mind during the early stages.

5. Summary

A *crystal* is made up by the indefinite repetition of a small three-dimensional unit, the *unit cell*, consisting of a small number of chemical molecules. It generally possesses symmetry elements, and the particular set of them which is present is known as the *space group* of the crystal. (Similarly for finite, nonrepeating objects exhibiting symmetry the set of symmetry elements is called a *point group*.)

X-rays are scattered by the electron density of the crystal. The diffraction pattern can be thought of as a regular three-dimensional array of spots, known as the *reciprocal lattice*. Each X-ray spot corresponds to one imaginary wave of a wave analysis, or *Fourier analysis*, of the electron density. Its position in the reciprocal lattice shows both the wavelength (or *spacing*) and the direction of the wave. Its intensity is related to the amplitude of the wave. Its phase is *not* given by the X-ray data. Therefore one cannot deduce the structure directly from the X-ray pattern, except in very simple cases; but, given the structure, one can always calculate the pattern.

What *can* we learn directly from the X-ray pattern of a crystal? The dimensions of the unit cell can be directly calculated from the dimensions of the reciprocal lattice. The symmetry of the crystal is closely related to the symmetry of the reciprocal lattice, and for protein crystals one can almost always deduce the space-group from a study of the X-ray pictures. Hence, knowing the volume of the unit cell, the size of the *asymmetric unit* can be calculated.

Powder patterns are familiar from industrial practice, and are obtained by passing a beam of X-rays through a crystalline powder. They can be thought of as the superposition of a large number of single crystal pictures of crystals in every possible orientation relative to the X-ray beam. Their characteristic feature is a set of concentric and sharp but continuous rings of blackening; the radii of these rings correspond to the spacings of the principal lattice planes in the crystal. One can think of them as generated by rotating the reciprocal lattice about all possible axes through its origin, and taking a central section of the resulting set of concentric spheres.

A *fiber* is usually a collection of small crystallites, whose "fiber axis" is nearly parallel to the length of the fiber. X-ray pictures of fibers are generally more confused and less perfect than those of single crystals, because the structure of most fibers is only partly ordered, so that some of the X-ray intensity is thrown into regions of diffuse scattering and not into discrete spots.

Fibers often possess nonintegral screw axes of symmetry, and the presence of these can often be deduced by inspection from the X-ray photo-

graph. Where adequate stereochemical information has been made available by studies of the structures of small molecules, it is sometimes possible to guess the structure of a fiber by careful model building, using accurate scale models.

X-ray diffraction is only really useful for studying that part of a structure which repeats regularly in space. By X-ray techniques it is easy to show that two structures are similar, but very difficult to show that they are identical, at any rate when the molecules are large.

The importance of symmetry, whether in a crystal, in a fiber, or in a virus, is that it allows the same subunits to be used in identical environments, repeatedly in the same structure. Presumably for reasons of economy in manufacture, Nature is addicted to the mass production of identical small units for building up large constructions. These tend to aggregate in a symmetrical manner which can be "seen" by X-rays. This is why X-rays are useful in studying biological structures. And this is why symmetry is the most important of all crystallographic ideas for biochemists.

III. Fibrous Proteins and Synthetic Polypeptides

In this section we shall give a brief account of recent work on the small-scale structure of fibrous proteins and synthetic polypeptides. The latter serve as model structures, simpler than naturally occurring materials because they can if desired have uniform side chains; and they have provided some of the most important clues to the configurations of collagens and keratins. A more detailed account of X-ray studies of fibrous proteins up to 1954 has been published by Kendrew (1954b), while the most recent advances have been reported by Kendrew and Perutz (1957). As stated in the introduction we shall not discuss in this review the *large*-scale structure of fibrous proteins.

1. Synthetic Polypeptides (and Silk)

The polypeptides which have proved most useful from the present point of view are those in which all the side chains are identical, though random copolymers having two or more types of side chain have also been synthesized. The degree of polymerization is generally fairly high—several hundred residues would be a typical value—so that the molecules are genuinely "fibrous." Oriented films or fibers can be produced by various simple techniques and these have occasionally given astonishingly good X-ray fiber diagrams.

It was discovered early that synthetic polypeptides form two main types of structure, known as the α- and the β- forms because they are analogous to the α- and β-forms of keratin. By appropriate choice of solvent either one or the other can be precipitated from solution at will. Thus *m*-cresol usually gives the α-form, while the β-form is precipitated from formic acid. The two forms give quite different X-ray patterns and they can also be distinguished by means of their infrared absorption spectra (as shown by the extensive studies of Elliott and his coworkers, 1956; see the review by Doty and Geiduschek, 1953).

a. α-Polypeptides. The best X-ray photographs of polypeptides in the α-form have been obtained from poly-L-alanine (Bamford *et al.*, 1954; Brown and Trotter, 1956) and from poly-γ-methyl-L-glutamate (Bamford *et al.*, 1952, 1953). These photographs have very characteristic features, and so far all polypeptides in the α-form have given similar X-ray patterns, though with varying degrees of perfection. The most detailed studies of them are those carried out by Bamford and his colleagues at Messrs. Courtaulds Ltd (Bamford *et al.*, 1956) and described in their recent book.

The main features are a strong meridional reflection of 1.5 A., discovered by Perutz (1951); a strong "layer line" of reflections with layer line spacing 5.4 A.; together with a strong reflection, spacing about 10 A. (depending on the side chain), on the equator.

It now seems certain that the configuration of the α-polypeptides is based on the α-helix of Pauling *et al.*, (1951). This is a folded configuration of the main chain; the positions of the atoms in the side chains beyond C_β are not specified by it. A diagram of the α-helix is given in Fig. 7. The polypeptide chain backbone follows an approximately helical path having a pitch of 5.4 A. and containing about 3.6 amino acid residues per turn. The translation per residue in the fiber axis direction is thus 5.4/3.6 = 1.5 A. The C_α carbon atoms, to which the side chains are attached, are all at a radius of 2.3 A. The whole structure is held together by hydrogen bonds running from the NH of one peptide group to the CO of another peptide group on the next turn of the helix.

The arguments in favor of the α-helix have already been rather fully set out elsewhere (Crick, 1954) and will be only very briefly recapitulated here. From the X-ray pattern it is possible to deduce unambiguously the parameters of the nonintegral screw axis: these are a rotation of about 100° and a translation of 1.5 A. From the density of the specimen and the dimensions of the unit cell it can be shown that the asymmetric unit consists of a single amino acid residue. The positions of the strong reflections, together with the fact that α-polypeptides can be stretched into a β-form, show that there is only one polypeptide chain per lattice point, rather than two or more intertwined. Only two structures can be built to this specification: one, the α-helix, has the same parameters as those observed; the other (described by Bamford *et al.*, 1952) is very much less satisfactory stereochemically.

A quite different approach is to build models without assuming any particular screw axis. It can be shown that if a polypeptide chain is to

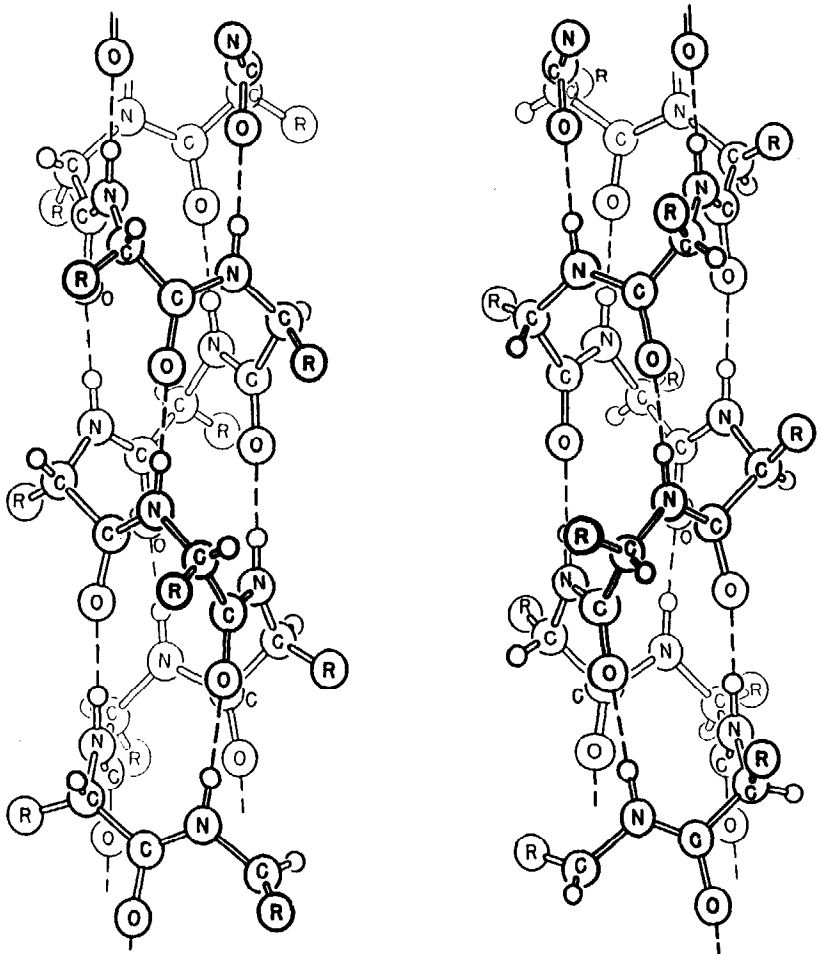

FIG. 7. Drawings of the left-handed and right-handed α-helices. The R and H groups on the α-carbon atom are in the correct position corresponding to the known configuration of the L-amino acids in proteins. (L. Pauling and R. B. Corey, unpublished drawings.)

be folded helically and stabilized by internal hydrogen bonds, as the infrared evidence suggests (Ambrose and Elliott, 1951), only a few simple arrangements are stereochemically possible. Moreover these can be enumerated systematically so that one can be certain that none have been overlooked.

All of them have been built by Donohue (1953) and arranged in a stereochemical order of merit. By any criteria the α-helix is the best, though one or two of the others cannot be totally excluded.

The strength of the case for the α-helix lies in the fact that both of these approaches give the same answer.

As originally described the α-helix was really *two* structures, since its backbone could follow either a right-handed or a left-handed helix. These are mirror images of each other; but there are two possible ways of adding side chains to the C_α carbon atoms of each, giving in all four structures, of which two are mirror images of the other two. Thus if we confine ourselves to L-polypeptides there are two possible structures, one with a right-handed helix, the other with a left-handed helix, but not mirror images of one another.

Until recently it was not known which of these two was more stable, but it now seems likely that the right-handed one is the more common for L-polypeptides. This had been suggested much earlier on structural grounds (Huggins, 1952); the newer evidence comes in part from studies on optical rotation, both experimental (Elliott *et al.*,1956; Yang and Doty, 1957) and theoretical (Moffitt, 1956 a,b; Fitts and Kirkwood, 1956a,b). In addition, a critical reconsideration of the X-ray data for poly-L-alanine (Elliott and Malcolm, 1956a) has shown that the agreement between calculated and observed X-ray intensities is greatly improved if the assumption is made that the chains are polarized at random either upward or downward in the structure; and that if this is done the right-handed α-helix fits the data much better than the left-handed.

There is a need for still more detailed comparisons between observed and calculated data for α-polypeptides, and these should make it possible to refine the structure even further. The presence of "forbidden" reflections, albeit weak, on the meridian of the poly-L-alanine pattern indicates that the α-helix must be slightly distorted in the solid state, probably owing to the mutual interference of neighboring chains. There must also be distortion in mixed DL-copolymers, since these give a 1.5 A. reflection whose spacing is slightly less than usual, and abnormally broad infrared absorption bands (Bamford *et al.*, personal communication). Model building suggests that in this case the distortion is due to occasional steric hindrance between C_β atoms of side chains (of differing hands) belonging to residues on adjacent turns of the same helix (Crick, 1953b).

There is now considerable evidence that the α-helix exists in solution in certain solvents, such as *m*-cresol, dimethyl-formamide, and chloroform-formamide, provided that the polymer be long enough (say 100 residues), since such solutions behave as if they contained rigid rods in which each residue occupies 1.5 A. of length (Doty *et al.*, 1956). Mixed DL-polymers

also form a helical configuration, but it is less stable than that of the pure D- or L-material. Thus it is clear that each enantiomorph of the monomer prefers its natural sense of helix (Doty and Lundberg, 1956; Elliott *et al.*, 1956). The stability of the α-helix in solution, in various solvents, has been discussed theoretically by Schellman (1955).

b. β-Polypeptides and Silk. It has been known for many years that in the so-called β-configuration of keratin and synthetic polypeptides, and also in silk whose X-ray pattern shows it to be a close relative, the polypeptide must be very nearly fully extended. A fully extended chain has a twofold screw axis, and repeats after two residues in a distance of 7.3 A. The observed repeat in all known β-structures is less than this—usually between 6.6 A. and 7.0 A.—so the chains must be somewhat puckered. In general plan the features of a β-structure follow from the disposition of the hydrogen bonds. The hydrogen bonding groups (CO and NH) of a single extended chain all lie in a plane, or nearly so, and project in a direction roughly perpendicular to the chain direction; it follows that the polypeptide chains can easily be hydrogen bonded into infinite plane sheets. The side chains project alternately on either side of the sheet. Neighboring sheets must be held together by bonds of various types between opposed side chains.

The main difficulty in making this general plan more precise is to know the relative directions of neighboring chains in the same sheet. Pauling and Corey (1953b) have described two possible regular arrangements: in the "parallel pleated sheet" all the chains in one sheet have the same direction, while in the "anti-parallel pleated sheet" alternate chains have opposite directions (see Fig. 8). They claim that if the structures are built so as to conform to the best values of bond dimensions the former arrangement gives a repeat of 6.5 A. and the latter a repeat of 7.0 A. It has not been rigorously established, however, that the two models can be distinguished merely by observing the exact value of the repeat, and it seems quite as likely that in the synthetic polypeptides the directions of the chains are randomly in one direction or the other (see Brown and Trotter, 1956).

We shall discuss silk only briefly: for a more extended account of recent work see Kendrew and Perutz (1957). For the silk of *Bombyx mori*, which contains 44% glycine residues, Marsh *et al.* (1955a,b) and also Warwicker (1954) have suggested a structure based on the antiparallel pleated sheet, in which it is supposed that every alternate residue along the chains is glycine; in consequence glycine residues all project (or rather, since their "side chains" consist merely of hydrogen atoms, fail to project) on one side of a given sheet. Two such sheets are packed back to back with their glycine sides together; and the whole structure is supposed to be

built up of pairs of such sheets (see Fig. 9). Chemical evidence on the amino acid sequence supports these ideas, and it seems very likely that

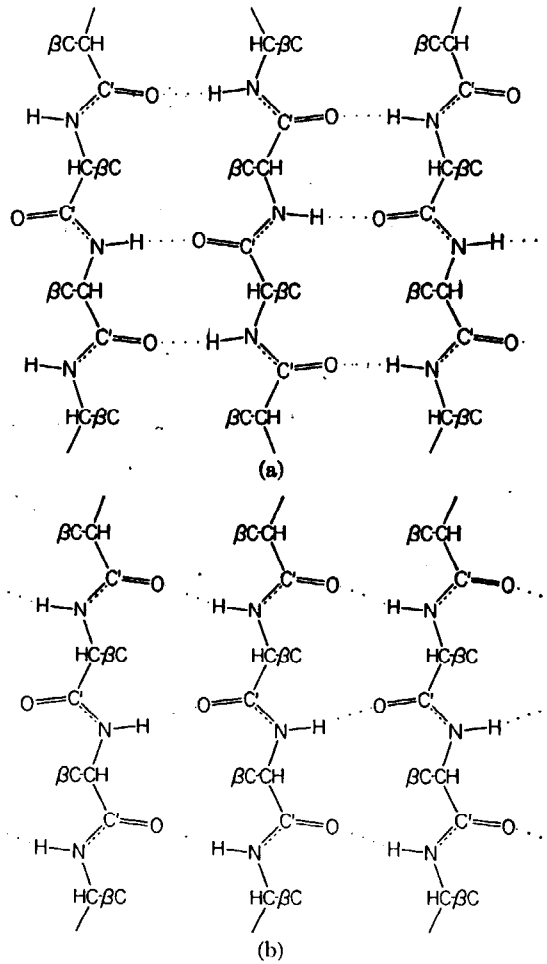

FIG. 8. Diagrammatic representations of the two pleated sheet structures proposed by Pauling and Corey (1951c and 1953b). (a) The anti-parallel pleated sheet, with chains running alternately up and down. (b) The parallel pleated sheet, with all chains running in the same direction.

much of the structure does possess the double-sheet structure. But the details, for example the direction of run of the chains and the interpretation of the longer equatorial spacings (Marsh *et al.*, 1955a), seem to us less certain.

In the more uncommon Tussah silk only 27 % of the residues are glycine so the arrangement must be somewhat different. Marsh *et al.* (1955c)

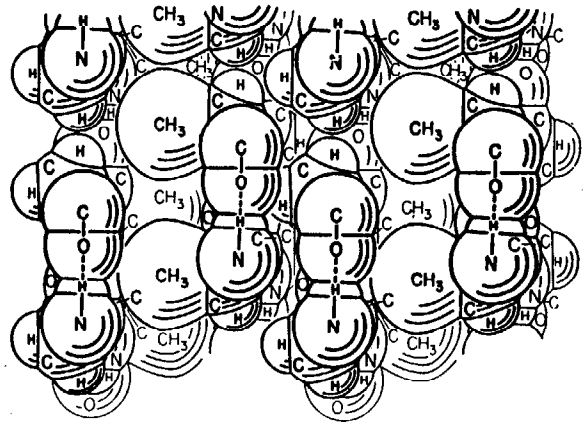

FIG. 9. The basic structure proposed by Marsh *et al.* (1955a) for the silk fibroin of *Bombyx mori*. The figure shows the view looking down the fiber axis, so that the polypeptide chains are running toward the reader. The sheets of polypeptide chains cut the figure in vertical lines. Notice two sheets close together, back-to-back, in the center of the figure, with alanine side-chains on the *outside* of the pair of sheets.

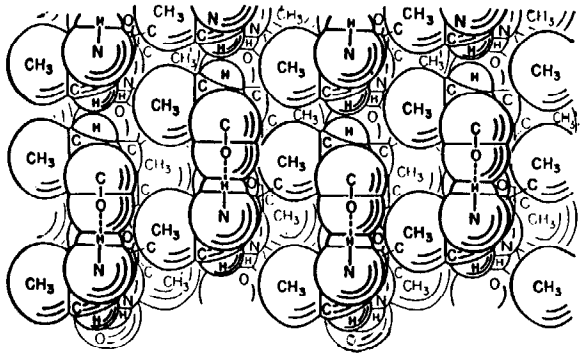

FIG. 10. The basic structure proposed by Marsh *et al.* (1955b) for Tussah silk, and also for the β-form of poly-L-alanine. Again the view is down the fiber axis. Notice that in contrast to Fig. 9 sheets of polypeptide chains do *not* occur in pairs but are equally spaced.

have suggested a simple structure, also based on the antiparallel pleated sheet; here it supposed that the glycines are arranged at random so that the two sides of any sheet are equivalent and all sheets pack at the same distance from one another, i.e. as singlets rather than doublets (see Fig. 10).

They propose a similar structure for the β-form of poly-L-alanine whose X-ray picture is remarkably similar (Bamford *et al.*, 1954; Brown and Trotter, 1956).

Finally it should be noted that by using appropriate solvents "soluble silk" can be made to take up the α-configuration (Ambrose *et al.*, 1951; Elliott and Malcolm, 1956b).

c. Polyproline. We now come to two materials which fall outside the classification of α- and β-structures. The first of these is poly-L-proline, which is interesting because of its relationship to collagen and because, having no NH group, it is incapable of donating hydrogen atoms for hydrogen bond formation. Cowan and McGavin (1955a,b) have studied its X-ray diffraction pattern, using material prepared by Katchalski. The X-ray pattern can be indexed in terms of a relatively simple unit cell of space group $P3_2$ and dimensions $a = 6.62$ A., $c = 9.36$ A.; in other words with a threefold screw axis. The asymmetric unit contains one residue, making the distance per residue in the fiber axis direction 3.1 A., a value which indicates that the polypeptide chain must be somewhat folded.

Model building shows that, if the peptide group is both *trans-* and planar, only a very limited number of configurations is at all possible, owing to the severe restrictions imposed by the steric hindrance between neighboring residues and by the fact that there is only one bond per residue about which rotation can take place. Only one of these configurations (see Fig. 11) has a triad symmetry axis. These considerations establish the general nature of the structure, although at the time of writing neither the exact details of the configuration nor the position of the molecule in the unit cell have been deduced unambiguously from the X-ray data—probably because, once again, the chains are running up and down at random in the structure.

There is no reason to suppose that the integral threefold axis is an especially favored configuration for the polypeptide chain. It probably arises in this case because of strong van der Waals' interactions between neighboring chains, which discourage the formation of a nonintegral screw (see p. 149).

d. Polyglycine. Polyglycine can be precipitated from solvents in two different forms having different X-ray patterns. That of polyglycine I is a typical β-pattern; but polyglycine II gives a new kind of pattern not hitherto obtained from any other material, although so far oriented specimens have not been obtained and the only photographs available are powder patterns (Meyer and Go, 1934; Bamford *et al.*, 1955).

Polyglycine II is of interest because of its relationship to collagen (see p. 168). It is prepared by precipitation from aqueous solutions in the presence of salts such as lithium bromide or calcium chloride. The struc-

ture turns out to be based on an integral threefold screw axis; the powder diagram can be indexed in terms of a trigonal unit cell, space group $P3_1$

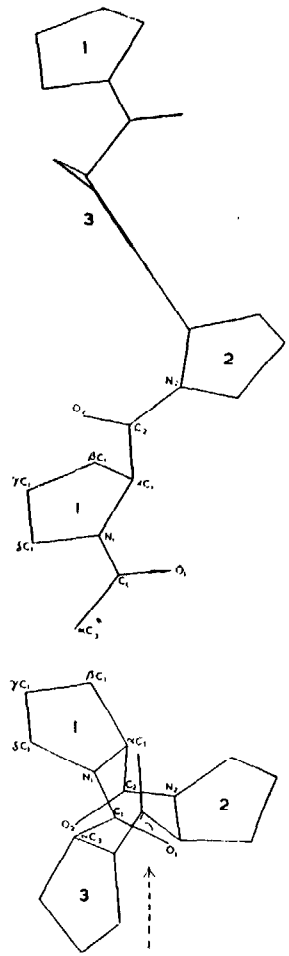

FIG. 11. A single chain of the model proposed for poly-L-proline, seen in projection, below, along the threefold screw axis, and, above, perpendicular to this axis and along the direction indicated by the arrow. (Cowan and McGavin, 1955b.)

or $P3_2$, $a = 4.8$ A., $c = 9.3$ A., one residue per asymmetric unit. The configuration proposed for polyglycine II by Crick and Rich (1955) (see Fig. 12) has a backbone configuration very similar to that of poly-L-proline. But in this substance, unlike the latter, the peptide groups all contain hydrogen atoms suitable for hydrogen bond formation, and in

the proposed structure neighboring chains are joined together to form an infinite three-dimensional network by three sets of hydrogen bonds running perpendicular to the fiber axis. The diffraction data are not sufficiently detailed to enable a decision to be made whether all the chains run in the

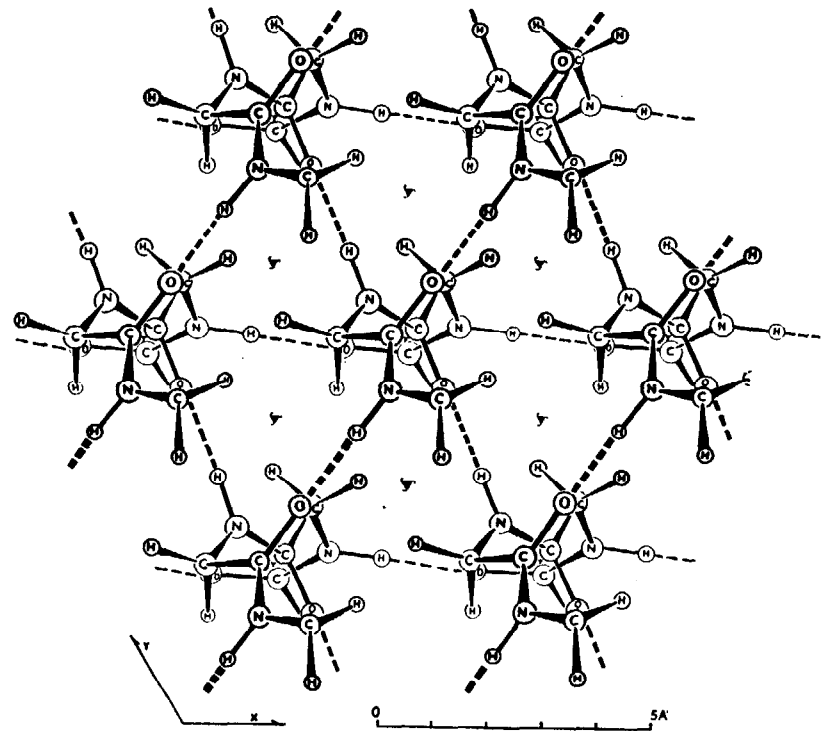

FIG. 12. The basic structure proposed for polyglycine II. A projection down the threefold screw axis, showing seven chains. Hydrogen bonds, drawn as dashed lines, run in a number of directions linking neighboring chains together. (Crick and Rich, 1955.)

same direction or whether they run randomly up and down; either arrangement would lead to a stereochemically plausible structure.

Further confirmation of the threefold character of the structure comes from the observations of Meggy and Sikorski (1956), who have found hexagonal crystals of polyglycine II in electron micrographs.

2. Fibrous Proteins

a. The α-Keratin Pattern. Hair epidermis, porcupine quill, myosin, tropomyosin, fibrinogen, and other naturally occurring materials give

diffraction patterns resembling one another in broad features, and known as alpha-patterns (for details see Kendrew, 1954a). The most striking of these features are strong meridional reflections with spacings of 5.1 A. (Astbury and Woods, 1930) and 1.5A. (Perutz, 1951). The latter strongly suggests that the structure is based on the α-helix. However, an array of parallel α-helixes would not give a 5.1 A. meridional reflection, but as pointed out by Crick (1952, 1953a) and by Pauling and Corey (1953a) a

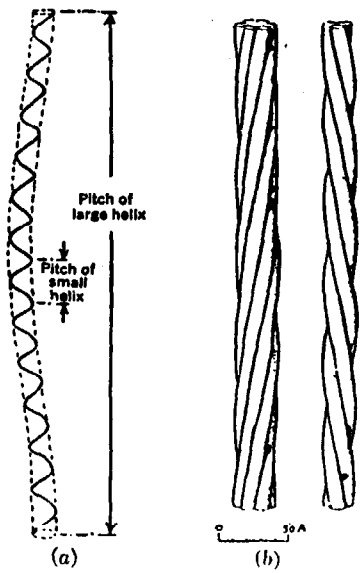

FIG. 13. To illustrate the general idea of a "coiled-coil" or "compound helix." The figure on the left shows a single polypeptide chain. The small helix is supposed to be an α-helix whose axis has been distorted so that it follows a larger, more gradual helix. The figure on the right shows two possible ways of combining helices into ropes. (After Pauling and Corey, 1953a.)

system of α-helices twisted together into coiled coils is capable of explaining both the meridional reflections.

It seems probable that this suggestion is correct in principle, but the details are still very uncertain. Pauling and Corey (1953b) proposed a complicated structure based on a 7-stranded rope, composed of a central straight α-helix with six others twisting slowly around it (see Fig. 13), together with additional interstitial α-helices; and they made the suggestion that the super coiling might be produced by a repeating sequence of residues. Crick (1953a) tentatively proposed two simple models—the double rope and the triple rope—which could be derived from simple

packing considerations: for reasons of symmetry two right-handed α-helices might be expected to pack together, not parallel, but at an angle of 20° to one another, when the side chains of one fit into the spaces between the side chains of the other. By a slight deformation this would yield a structure resembling a piece of twin lighting cable; the triple rope is similar. Recently Lang (1956a,b) has shown that this kind of structure would probably give an X-ray pattern simpler than that observed, although his argument is not entirely rigorous, since he made no allowance for side chains.

Not only are the details of the configuration unknown, but it seems likely that they may be different in different materials giving the α-keratin pattern. Tropomyosin, for example, with no proline and little cystine, and a molecular width corresponding to only two polypeptide chains, is unlikely to have precisely the same structure as porcupine quill, which contains large amounts of both proline and cystine and gives an X-ray pattern of considerable complexity.

In spite of these reservations it seems almost certain that a substantial part of these proteins is folded into the α-helix configuration, so we may be reasonably confident that the α-helix is not restricted to synthetic polypeptides but can also occur in genuine proteins.

b. The β-Keratin Pattern. It was shown many years ago by Astbury and his colleagues (1930, 1931, and 1933) that when hair is stretched its X-ray diagram changes from what is now called the α-pattern to a radically different one called the β-pattern; he concluded that this change reflected a change in the configuration of the polypeptide chain from a folded form to one which is almost fully extended. This interpretation is still considered to be correct; but the details of the β-configuration have eluded discovery, although its general nature is not in doubt. What we have said above (p. 158) in connection with β-polypeptides and silk applies also to the other β-proteins, which include the stretched forms of many of the proteins we have listed above as α-proteins, as well as feather keratin, which exists only in what is presumably a β-configuration.

The most important features of the X-ray pattern of β-keratin are equatorial reflections of spacing 9.7 and 4.65 A., and a meridional reflection of spacing 3.33 A.; there is no reflection of spacing 1.5 A., but instead one of 1.1 A. Pauling and Corey (1953b) have suggested that the structure is essentially a parallel pleated sheet, with repeating unit 6.5 A., in contradistinction to β-polyalanine and silk to which, it will be remembered, they have attributed the antiparallel pleated sheet. In our view the experimental evidence does not yet permit the deduction of so precise a model; but in general terms it does seem likely that the structure consists of pleated sheets or something very like them. It is to be hoped that more

definite conclusions can be reached by working on the few β-proteins which give diffraction patterns rich in detail: among these is feather keratin, on which a preliminary note has been published by Krimm and Schor (1956).

One of the problems still to be solved about the structure of keratin is the exact nature of the α-β transformation. This process is reversible and takes place under relatively mild conditions. Keratin contains a very large number of S—S bridges, and the chemical evidence suggests that these are not ruptured during extension. It is not easy to see how a process which, we must presume, involves the pulling out of (possibly intertwined) helices into pleated sheets, could leave so many interchain bridges intact; the suggestion that all the S—S bridges are *intra*-chain also raises formidable stereochemical difficulties. Nor is it clear how all the side chains can readjust themselves easily, since in some cases one would expect them to rotate through large angles during the extension.

c. Collagen. In this review we shall be concerned only with the structure of collagen at the atomic level, although it also exhibits features of great interest at a higher level which can be studied in the electron microscope. For a recent discussion of the latter see Schmitt *et al.* (1955).

Collagen has been studied by X-rays for many years. As Anfinsen and Redfield have said in the review to which we have already referred (1956), "perhaps for no other protein has such a multitude of structures been proposed, or, to use a term more common among X-ray crystallographers, 'discovered'." It must be admitted that the jibe was not unjustified; the structure was in fact unknown until recently, when several groups of workers suggested essentially similar solutions. These suggestions have radically altered the situation, which now is that the structure is almost certainly "discovered" in a final sense of the term. It is unlikely that the recent models will require modification except in detail. There have been several reviews of the earlier efforts (Bear, 1952; Kendrew, 1954a), which need not be described here.

It is well known that collagen has an unusual amino acid composition (see Tristram, 1953). Its main peculiarities are the high glycine content (just over one-third of the residues are glycines); the presence of hydroxyproline and hydroxylysine, amino acids which occur in no proteins other than collagen and its near relations; and the large amounts of proline and hydroxyproline, which together make up about 22% of the residues of beef collagen. There have been two important studies of the amino acid sequence in collagen. The first, by Schroeder *et al.* (1954), showed that the sequence Pro-Gly was rare and Gly-Hypro absent, whereas Gly-Pro and Hypro-Gly were common. The provisional conclusion, that Gly-Pro-Hypro-Gly might be a common sequence in collagen, was confirmed by Kroner *et al.* (1955), who identified this tetrapeptide among their hydrolysis products, as well as the tripeptide Gly-Pro-Hypro. It seems very probable

that the conclusion may be accepted, in spite of the rather low yields obtained in both studies.

The X-ray pattern of collagen is of a type given by no other protein.

Its main features are a strong meridional arc of spacing 2.86 A. and near-meridional spots with spacings about 4 and 10 A. There are also equatorial reflections, the principal among which is humidity-sensitive, having a spacing of 10.4 A. in dry and up to 17 A. in wet collagen. Finally there is a diffuse patch on the equator in the 4½ A. region, especially strong in the dry material (see Fig. 4).

Certain electron micrographs (Schmitt *et al.*, 1942; Mustacchi, 1951) have suggested that collagen fibers may be able to stretch by large amounts (up to several hundred per cent); but no one has been able to reproduce this phenomenon except under electron bombardment, so it seems probable that it is an artifact. There is no doubt, however, that collagen can be stretched reversibly by *small* amounts (up to about 10%) and that during stretching there is an increase in the spacing of the principal meridional reflection, normally 2.86 A. This effect was discovered by Cowan *et al.* (1953), who found that it was accompanied by a considerable improvement in the definition of the X-ray pattern, and thus made an important technical advance. They suggested (1953), as did Cohen and Bear (1953), that the structure was based on a nonintegral helix. There is now general agreement with this view and that the approximate parameters of the screw axis are a rotation of 108° and a translation of 2.86 A.

To be more correct, the structure might have an n-fold *rotation* axis, parallel to the fiber axis, in addition to the screw, whose parameters would then be $108°/n$ and 2.86 A. Consideration of the distribution of strong intensities in the diffraction pattern, and of the probable mean radius of the helix, makes it very likely that in fact $n = 1$ (Cowan *et al.*, 1955).

The net-diagram implied by this screw symmetry is shown in Fig. 5, but it cannot tell us which way the polypeptide chains run, nor how many of them there are, even though there is independent evidence that the number of amino acid residues per asymmetric unit is three (this follows from a consideration of the density of the structure). Various possibilities are shown in Fig. 5. In fact there is evidence from studies of light scattering, etc., on collagen in solution (Boedtker and Doty, 1956), as well as from the model-building approach which we shall now discuss, that the number of chains in the helix is most probably three.

The structures recently proposed, which are all closely related though not identical, spring from an earlier suggestion by Ramachandran and Kartha (1954). Their first model (whose symmetry is *not* the same as that discussed above) consisted of three parallel polypeptide chains, joined by hydrogen bonds, not twined around a common axis but running

side by side in a compact group. Each chain had a threefold screw axis with a translation of 9.5 A. (containing three residues) in the fiber axial direction. The backbone configuration of these chains was, in fact, very similar to that subsequently established for polyproline and polyglycine II, whose axial repeats are almost the same. Ramachandran and Kartha later (1955) modified their structure by causing the three chains to twist slowly around each other, thus giving the model a nonintegral screw axis in conformity with the net-diagram discussed above.

The other structures suggested recently have all been of this form, but have differed in the way the three chains are linked together. Since the asymmetric unit contains three residues one can conceive of three different types of interchain hydrogen bond. Rich and Crick (1955) have shown by exhaustive model building that only one of the three types can be made *systematically*, between atoms of the polypeptide backbone: all structures with more than one type are stereochemically unsatisfactory. Moreover, there are only two ways of making a single set of hydrogen bonds, and these they have described as Structure I and Structure II. The same conclusion has been reached by Bear (1956), also from systematic model building, but to a slightly different set of postulates.

It will be remembered that in polyglycine II an infinite network of hexagonally arranged chains is linked together by hydrogen bonds (Fig. 12). We might imagine the collagen structure as derived by isolating a group of three chains from this infinite network. It can easily be shown by means of models that there are just two types of groups which can be isolated in this way, differing in the way their hydrogen bonds are arranged. One of these types corresponds to Structure I for collagen, the other to Structure II.

Structure II (Fig. 14) turned out to be much easier to bu ld than Structure I, in that it gave more acceptable values of bond dimensions and angles and of interatomic distances; also its diffraction pattern is in better agreement with the observed pattern (Ramachandran, 1956; Bear, 1956; Cowan *et al.*, 1955; Rich and Crick, unpublished). For stereochemical reasons Structure II will accommodate only the amino acid sequence —G—P_1—P_2—, repeated indefinitely; G must be glycine, while P_1 and P_2 could be any residues, including proline and hydroxyproline. The amino acid sequence data to which we have already referred indicate that in fact all the hydroxyproline must be at P_2. This site is located far from the axis of the structure, and thus in Structure II the hydroxyl group of hydroxyproline cannot form hydrogen bonds with CO groups in the backbones of the same group of chains. It follows that if it is used for interchain linkages at all, it must serve to link together neighboring groups of chains, as suggested by Ramachandran and Kartha, rather than to link chains within one group, which was the case in the less satisfactory

Structure I (Rich and Crick, 1955). The data collected by Gustavson (1955), suggest that the thermal stability of collagen is greater the greater

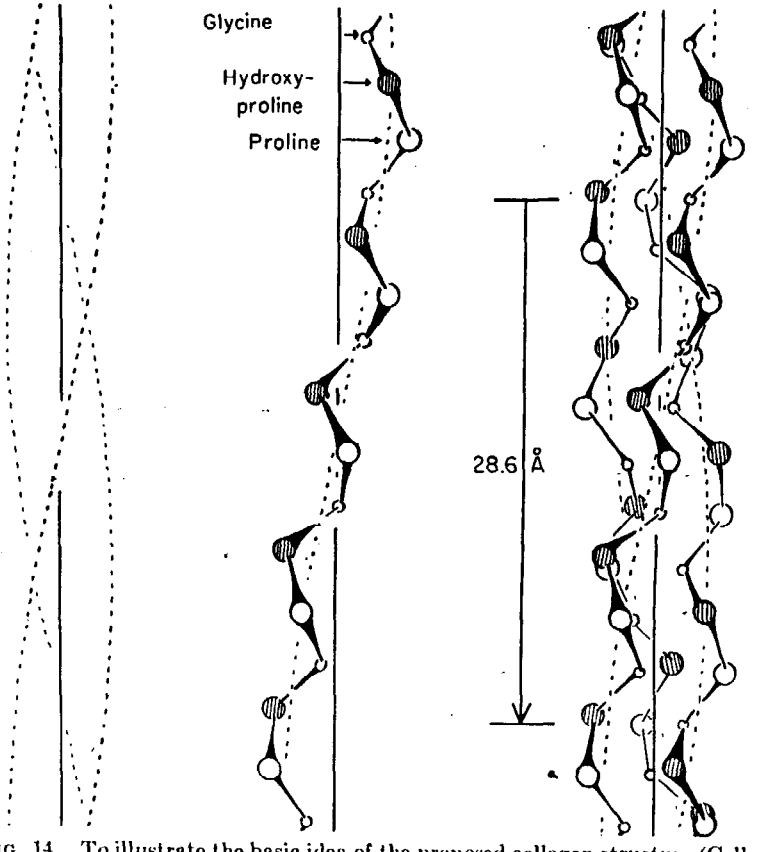

FIG. 14. To illustrate the basic idea of the proposed collagen structure (Collagen II of Rich and Crick, 1955). For clarity *only the C_α carbon atoms are shown.* The peptide groups connecting them are drawn simply as short straight lines. On the left the dotted lines show the general run of the three polypeptide chains about the fiber axis (full line). In the middle *one* of the three chains is shown, to illustrate how it coils round the dotted line. On the right all three chains are included. The small circles show the sites which must be glycine. The large circles show where proline and hydroxyproline (shaded) are mainly found. Note the repeating sequence of sites.

its content of hydroxyproline; but as yet there is no chemical evidence whether the bonds it forms are with a group of three chains, or between groups, or both.

It should be noted that in these structures relatively few hydrogen

bonds are made between backbone atoms. There seems to be no intrinsic objection to this, however; indeed it may be that the solution of the structure has been delayed by an overemphasis on backbone-backbone hydrogen bonds. It cannot be said that there is yet general agreement that Rich and Crick's Structure II is correct. There *is*, however, general agreement that all other structures so far proposed are unsatisfactory, and that Structure II is the best suggestion yet. In our opinion it is likely that it will turn out to be correct. Nevertheless it is necessary to add a note of caution to the effect that different parts of the collagen molecule may have different configurations. It is well known that collagen fibers have a banded structure which can be seen in great detail in electron micrographs, and that the bands differ in certain respects from the interbands. Moreover collagen has to be stretched in order to give a good diffraction pattern; it may be that the effect of stretching is to alter the configuration of part of the fiber. It is not impossible, in fact, that in unstretched collagen part of the chain has a different configuration, Structure I for example. If it were shown that the collagen molecule is inhomogeneous in some such sense as this, the force of some of the arguments used to deduce the structure would naturally be weakened.

IV. Crystalline Proteins.

More work has been done to determine the structures of the globular proteins by means of X-rays than has been done in any other area of the field, and with fewer results. By and large, globular proteins are metabolically active, and fibrous proteins are not. From the biochemist's point of view, therefore, any results obtained with globular proteins should be the most interesting of all. This branch of protein X-ray studies has in fact just reached a critical point. For the first time there is a real prospect of getting definite and incontrovertible results. None to speak of have yet been published—the achievements so far are spectacular from the technical standpoint, but not from the point of view of the interested outside observer—but there is now for the first time a real promise for the immediate future. The transformation of the field is largely a consequence of the successful application, by Perutz and his colleagues, of the method of isomorphous replacement to a protein crystal. We shall speak of this in its place; in the meantime we must make a preliminary survey of some basic facts about protein crystals.

1. The Nature of Protein Crystals

The main difference between protein crystals and the crystals of much smaller organic molecules is that they contain a considerable quantity of solvent, actually within each unit cell. Typically half the volume of

the crystal will be water (or, more often, the salt solution with which the crystal is in equilibrium). If such a crystal is removed from its mother liquor and exposed to the air, water is lost and the crystal can be seen to shrink somewhat; its optical properties usually deteriorate at the same time. X-ray measurements would show that the visible shrinkage is a consequence of the shrinkage of the unit cell itself. In this condition a crystal is conventionally described as "dry," in contradistinction to the original "wet" crystal—though in fact it can be dried still further if placed in a desiccator.

All the evidence suggests that most of the water in the crystal is in a "liquid" state—that is to say, it has no regular structure like ice or like the hydrated layer around an ion; and it is permeable to small ions. Thus considerable amounts of salts can often be diffused into a crystal without changing the dimensions of the unit cell, and indeed, since many proteins are crystallized by "salting out," the salt concentration in the liquid inside the crystal may reach several moles per liter. Proteins such as ribonuclease, which are crystallized from strong solutions of organic solvents, exhibit similar behavior in that the cell dimensions hardly change when the organic solvent is changed, a typical alteration (for ribonuclease) being 0.1 A. in 30 A. In all these cases, the fact that ions or other small molecules have gone right into each unit cell can be demonstrated in several ways. For example, if sodium dithionite is diffused into a crystal of methemoglobin the spectrum of the protein can be directly observed to change from that of ferrihemoglobin to that of ferrohemoglobin, as the process of diffusion takes place. Again, changes in the low order X-ray reflections (that is, the reflections of long spacing near the center of the photograph) show clearly that these small molecules have penetrated the unit cell. This is demonstrated in Fig. 15 which shows the reflections of finback whale myoglobin in four different salt solutions. Note that while the inner reflections alter dramatically, the outer ones are unchanged. This shows that whereas the *fine structure* of the contents of the unit cell is unaltered, the *general distribution* of electron density, as seen at low resolution, has altered greatly; and the effect is completely explained by supposing that the electron density of the structureless but extensive (and, in regard to their boundaries, somewhat ill-defined) regions containing mother liquor has been stepped up or down, while that of the protein molecules, with their precise and definite structure, has remained unchanged.

Still larger molecules, such as dyes or compounds like *p*-chloromercuribenzoate used for isomorphous replacement (page 183) can diffuse into the crystal, and once again the X-ray data show clearly that these molecules really go inside the unit cell and not merely between crystallites. In

favorable cases the extra molecules have no effect on the dimensions of the unit cell; but sometimes small changes in dimension do take place, and sometimes the crystals may become disordered or perhaps break up alto-

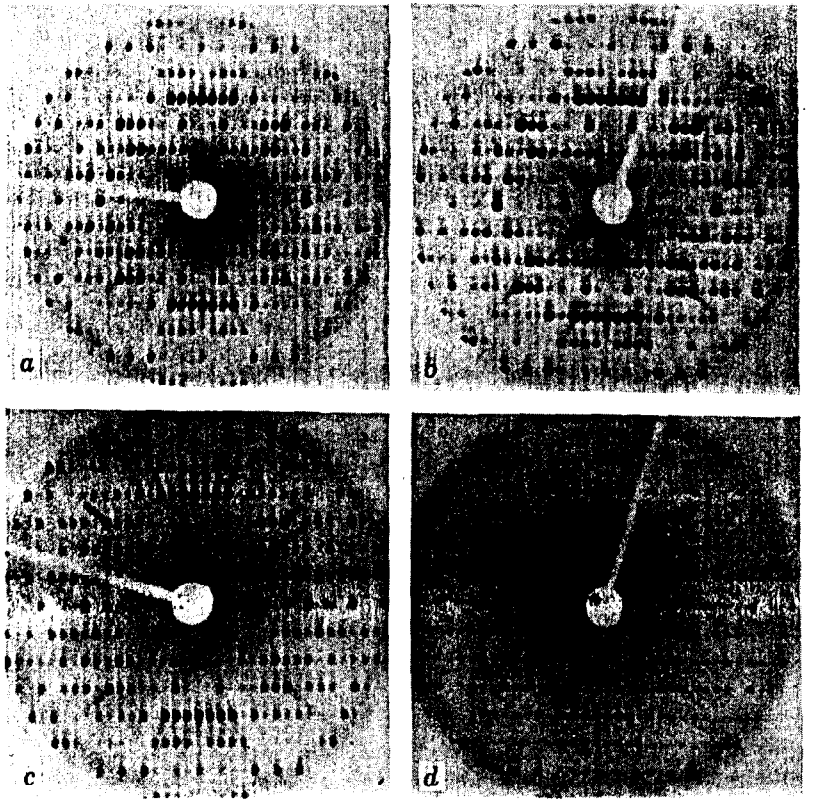

FIG. 15. To illustrate the penetration of salts *within* the unit cells of a protein crystal. The salt concentrations in the different suspension media were (a) 75% saturated $(NH_4)_2SO_4$, (b) saturated $(NH_4)_2SO_4$, (c) 4 *M* phosphate, and (d) 7 *M* phosphate (approximately). Notice that while the intensities of the outer spots remain the same, those of the inner ones change very considerably as the electron density of the medium is altered. (Finback whale myoglobin, [001] zone; Kendrew and Pauling, 1956.)

gether. Occasionally a radically new unit cell is formed, as when the dye iodophenol blue is added to ribonuclease (Magdoff and Crick, 1955b), or when *p*-iodophenyl hydroxylamine is crystallized with myoglobin.

A most interesting example of the same phenomenon is provided by oxy- and reduced hemoglobins, which normally crystallize in totally different space groups, suggesting that the process of oxygenation may involve an appreciable change in the shape of the molecule, possibly by altering the relative positions of the two subunits of which it is composed (see p. 175). It is noteworthy that in myoglobin, where the indications are that the molecule is *not* made up of subunits, no such phenomenon has been observed: oxy- and reduced myoglobins crystallize isomorphously.

These observations leave no room for doubt that some of the water (or other solvent) within the protein crystal is "liquid;" and the question arises whether it is *all* liquid. Only in horse hemoglobin has it been fully answered, by Perutz (1946), who measured the density of the crystals after they had been equilibrated in salt solutions of various concentrations. His results are in accordance with the assumption that part of the water is "bound" to the protein, that is to say, held in a rigid or pseudocrystalline arrangement, so that salt cannot diffuse into it; on the other hand, the rest of the water is continuous with the external medium and contains the same concentration of salt. The amount of "bound water" was found to be 30% of the protein (by weight). Whether this simple picture has any physical reality remains to be seen: but at least it summarizes the facts in a very compact way. On the other hand, despite earlier suggestions, the X-ray data clearly show that it is an oversimplification to conclude that the bound water consists of a uniform unimolecular layer covering an ellipsoidal protein molecule (Crick, 1953a).

The shrinkage of protein crystals can also be studied by X-rays. It is often found that well-defined shrinkage stages exist between the wet and dry extremes. These stages have been carefully studied in horse hemoglobin (Huxley and Kendrew, 1953). The cell dimensions change quite sharply as the humidity is varied at a fixed temperature. In this protein it is even possible to obtain an "expanded" stage by altering the pH. Shrinkage stages have also been reported for various myoglobins (Kendrew, 1950; Kendrew and Pauling, 1956; Kendrew and Parrish, 1956), and also for ribonuclease (Magdoff and Crick, 1955b). For the latter protein it has been shown that the wet lattice can be "strained" by what appear to be small humidity variations; that is, the cell dimensions can be altered by about 0.3 A. in 30 A. in an apparently continuous manner (Magdoff and Crick, 1955b). It is not known whether this is true of any other protein.

All these phenomena can be understood if we regard a protein as a large molecule of relatively fixed size and shape. It would be surprising if such molecules (as opposed to smaller and relatively more flexible organic molecules) were able to pack together without leaving considerable space between them. This space is naturally filled with water, or other solvent, as in many crystals of smaller organic molecules; but it is bigger, and there

is room for a larger number of solvent molecules, which therefore find it easier to retain their "liquid" state—in other words they distribute themselves over the rather large space in a random manner. The protein molecules presumably touch one another at a rather small number of specific points of contact; and at least some of these are changed abruptly when the crystal goes from one shrinkage stage to another, although minor humidity changes may strain the arrangement a little without causing large and discontinuous changes. As more and more water is removed the molecules pack down together as best they can, and the structure often becomes disordered. If the crystal is, finally, dried thoroughly most of the water comes out of the interstices and empty spaces are left between the closely packed protein molecules.

In some proteins, β-lactoglobulin for example, it has been reported (McMeekin *et al.*, 1954) that shrinkage is continuous; though of course the truth may be that even here shrinkage stages do exist, but that their dimensions are so similar that they elude detection.

The X-ray pattern of wet protein crystals usually extends to spacings of about 1½ or 2 A, the average diffracted intensity falling rapidly with increasing spacing in this region. In this respect protein crystals differ from crystals of ordinary organic molecules, which produce diffracted beams of much smaller spacing. The absence of fine detail in protein diffraction patterns sets a limit to the resolution of the structure we can hope to obtain even when X-ray methods reach their ultimate power, although in some small proteins it may just be possible to resolve individual atoms. Some protein crystals are better than others from this point of view; thus ribonuclease is particularly good, with spots extending out to about 1½ A. In general the smaller the protein the further out into reciprocal space its diffraction pattern extends.

Dry crystals are always more disordered than wet ones, and generally give few reflections with spacings less than 5 A, though there are exceptions (in both directions). Thus the diffraction patterns of wet crystals contain more information, and it is usual to study proteins wet rather than dry. To do so one must mount them in sealed capillaries, as thin as possible to minimize loss of X-rays, and containing a few drops of mother liquor to stabilize the humidity.

2. Direct Information.

In this section we shall describe the sort of information which can be obtained from the preliminary examination of a protein crystal. Most of it (except that discussed under *d*) can be got in only a few days.

a. Unit Cell and Space Group. In most cases the dimensions of the unit cell and the nature of the space group (i.e. the symmetry elements)

can be derived unambiguously from two or three suitably chosen X-ray photographs. Reference to the *International Tables of Crystallography* at once gives the number of asymmetric units in the unit cell; and from the volume of the latter it is simple to calculate the volume of the asymmetric unit. If the molecular weight of the protein is approximately known one can calculate the maximum number of molecules which the asymmetric unit can contain. To obtain the actual number one must estimate the relative proportions of protein and solvent in the crystal. This usually presents little difficulty since in general only an approximate estimate is required; in fact in almost all cases the proportion of solvent is 40–60%. The most usual number of molecules in the asymmetric unit is one, but two are found quite commonly, and larger numbers occasionally. It may even happen that the number is a fraction. Thus in the most common form of horse hemoglobin, whose space group is C2, the number is one-half, showing that the "molecule" found in solution, of molecular weight 67,000, must consist of two identical halves. In the crystal these halves are related by the dyad or twofold rotation axis of symmetry which in this space group relates two neighboring asymmetric units. It is most likely that the same is true of a hemoglobin molecule in solution (it will be realized from what has so far been said that the environment of a protein in a crystal is rather like its environment in solution). In conditions of extreme dilution or in presence of high concentrations of urea the horse hemoglobin molecule dissociates into two halves in solution.

The contrary proposition—that a molecule possessing internal symmetry must exhibit it in the crystal—is not necessarily true. Sometimes a protein with internal symmetry may crystallize in two different forms, in one of which the internal symmetry forms part of the symmetry of the cell, while in the other it is not revealed. Thus X-ray evidence alone cannot tell us the minimum structural unit of the protein (at least from the preliminary examination). Insulin, for example, which has a chemical molecular weight of 6000, has a crystallographic molecular weight of 12,000 in both its known crystal forms, and this is also the lowest value so far found in aqueous solution. It will be interesting to see how the two halves of the 12,000 molecule are related, but this we shall not discover without a full-scale analysis of the crystals. Recent work has shown that dissociation of the 12,000 unit into "monomers" of molecular weight 6000 is promoted by urea and guanidine (Kupke and Linderstrøm-Lang, 1954; Trautman, 1956); this suggests that hydrogen bonds play an important role in holding the two parts together.

b. Molecular Weight. In favorable cases it is possible to obtain a rather good value of the molecular weight of the asymmetric unit of the protein

by X-ray studies; as we have just indicated, this may be a multiple or a submultiple of the "molecular weight" found by other methods. As this subject has been reviewed elsewhere very recently (Crick, 1957) it will only be briefly alluded to here. In essence the method is to measure the volume of the asymmetric unit (by measuring the cell dimensions), and to calculate its weight by measuring the density of the crystal. To determine the molecular weight of the protein it is then only necessary to establish the *composition* of the asymmetric unit in terms of protein, solvent, and salt (if any). This is easiest when salt or organic solvent is absent—if salt is present there will be difficulties due to the fact that part of the water is "bound" and salt-free, so that the overall salt concentration in the internal medium is less than that in the external medium. Usually therefore one works with salt-free crystals if these are available. Under favorable circumstances the errors should not exceed 1–2%, and even a very rough estimate will usually be within 5–10%. Considering its accuracy, the method has been somewhat neglected in the past; partly perhaps because it requires collaboration between a protein chemist and crystallographer. In Table II we have collected some of the more recent and more accurate results obtained by this method.

TABLE II

Molecular Weights of Some Proteins, as Determined by X-Rays

Protein	Molecular weight	Determined by
Ribonuclease	13,400	Harker (1956)
Lysozyme	13,900 ± 600	Palmer *et al.* (1948)
α-Chymotrypsinogen	25,000 ± 800	Bluhm and Kendrew (1956)
β-Lactoglobulin	35,000 ± 400	Green *et al.* (1956)
Human serum albumin	65,200 ± 1,300	Low (1952)
Human mercaptalbumin	65,600 ± 700	Low (1952)

c. Identification and Identity. It might be thought that the X-ray diffraction pattern, being so intimately related to the structure of the protein producing it, could be used like a finger print for identification purposes. Unfortunately this is true only to a limited extent. The same protein, crystallized under slightly different conditions, may give various crystal forms with totally different space groups and diffraction patterns. Thus the fact that the X-ray pictures of two protein crystals are radically different does not mean that the proteins themselves are different. On the other hand, two proteins known to be different (though the differences are slight) may sometimes crystallize in the same unit cell, and give almost identical diffraction patterns. The reason why this is possible has already been discussed (see p. 152).

The various crystal forms of myoglobin provide some very good examples of this (Kendrew *et al.*, 1954). Thus the form known as Type A has been obtained from sperm whale, finback whale, blue whale, sei whale, lesser rorqual, and common porpoise; the crystals are isomorphous and the diffracted intensities very similar though not identical. Again, crystals of the form called Type C have been obtained from the horse, common seal, and gray seal. Nevertheless the myoglobins of the different species differ immunologically and (wherever analyses have been made) chemically, albeit slightly. Changes in the diffracted intensities, of the same order of magnitude as those found in these examples, can also be produced by simple chemical modification of the protein, as for example by converting CO-myoglobin to metmyoglobin. It seems very probable, by analogy, that the changes produced by varying the species are a consequence of a few variations in the side chains (cf. the species variations in insulin investigated by Brown *et al.*, 1955).

Thus while the X-ray pattern is not a safe guide to strict identity, it remains true that if two proteins from different sources give very similar unit cells and diffraction patterns, it is virtually certain that they have the same major structural features, and therefore that their amino acid sequences are closely related.

d. The Shape of Protein Molecules. In certain special cases it is possible to learn something about the shape of the protein from the dimensions and symmetry of the various unit cells in which it occurs. It is rare that straightforward deductions can be made, however, and the information obtained is not generally very precise, so we shall only touch on it briefly (for a more extended account see Kendrew, 1954a).

The cell dimensions put upper limits to the diameter of the molecules in certain directions, but the restrictions are not often severe enough to be interesting. If the same protein crystallizes in many different forms it may be possible to deduce a unique shape for the "equivalent ellipsoid" such that good close-packing is achieved in all the forms. The most fully worked out example of this approach is hemoglobin, and those interested in it should consult the original papers (Bragg and Perutz, 1952b; Bragg *et al.*, 1954).

A source of information which is more often profitable is the inmost region of reciprocal space—the reflections of very low order—especially when the electron density of the solvent is very different from that of the solvent. These reflections, which correspond to a view of the structure at very low resolution, depend on the general contrast between the protein molecule and the solvent, and very little on the internal structure of the protein. Alternatively, when the salt concentration inside the structure is high, one can measure the *changes* in X-ray intensity produced by *changes*

in salt concentration (see Fig. 15). By methods of this sort Bragg and Perutz (1952a) derived a shape for the molecule of horse hemoglobin which agreed well with that deduced from packing considerations: namely an ellipsoid with dimensions 71 × 53 × 53 A. consisting of hydrated protein. This shape, however, can only be regarded as a first rough approximation to the truth—the molecule is almost certainly more asymmetrical and more "knobbly" than an ellipsoid. In our view the method of isomorphous replacement offers a more general and more reliable method for discovering the shape of a protein; and its application for this purpose will be discussed on page 195.

3. The Patterson Function

The basic principle of the Patterson synthesis has already been mentioned. In the past it was, for lack of anything better, the main tool for the exploratory studies of protein crystallographers. The newer methods have reduced its importance and we shall refer to it only very briefly here. The subject has been dealt with rather fully by one of us (Kendrew, 1954a), and a simple explanation of the ideas involved in its application to protein crystals has been set out in an earlier article (Kendrew and Perutz, 1949).

It will be recalled that in this method the experimental data are manipulated mathematically without any assumptions about the structure being made. This treatment gives a map which shows not the structure itself but the relative positions of all possible pairs of atoms in the structure, all superposed. It can be shown that if a structure, even so complicated a one as a protein, possesses certain strong features, such as parallel "rods" of high electron density (e.g. polypeptide chains in suitable configuration), the Patterson synthesis will possess analogous features. The actual interpretation is controversial in almost all cases. It suffices to say that the Patterson approach has clearly demonstrated that the structures of the few proteins so far examined are not of extreme simplicity in the sense of consisting essentially of bundles of parallel straight polypeptide chains; on the other hand they are certainly not completely isotropic. Some proteins, such as myoglobin, show more obvious signs of regularity than do others, such as ribonuclease.

Another application of the Patterson synthesis is to obtain relative orientations of the same protein in different unit cells, by considering the relative orientation of the strong features of their Patterson syntheses. This can be a powerful method in favorable cases, especially if three-dimensional data are available—but the computation of three-dimensional Patterson syntheses is at best a very tedious business, and it is doubtful if the effort is well spent now that more powerful, though equally tedious,

methods of analysis are available. Again, it may be possible to obtain some knowledge of the relative positions of the molecules in the unit cell by looking for "pseudo-origins"—that is, for regions where the Patterson function appears to repeat within the unit cell. These methods have been used for ox hemoglobin (Crick, 1956) and for various types of myoglobin (Kendrew and Pauling, 1956; Kendrew and Parrish, 1956). But in all cases the results are suggestive rather than conclusive, and so far there has been no opportunity to check them by more certain methods.

The Patterson synthesis, then, will always be a powerful tool in the hands of the crystallographer, but for the present any results obtained by its use should be accepted with reserve. The use of the Patterson synthesis in the isomorphous replacement method (see p. 181) is in a different category, however.

4. Methods Involving Heavy Atoms

These methods, which involve the addition of heavy atoms to the protein molecule and studying the consequent alteration in the diffraction pattern of the crystals, are the only ones so far discovered which give any secure hope of solving the structure of proteins. For this reason, and because they are intimately connected with the chemistry of proteins, we shall describe them at length. There are two distinct methods, both of which have been used for a number of years in the study of small molecules. The first has not yet been applied to proteins, while the second was so used for the first time in 1953.

a. The Heavy Atom Method. The first is the Heavy Atom Method proper. This was used by Carlisle and Crowfoot (1945) in their determination of the structure of cholesterol iodide; and also by Crowfoot-Hodgkin and her collaborators (Hodgkin *et al.*, 1956) in the first stages of the study of Vitamin B_{12}. In this method a heavy atom is introduced into the molecule, sufficiently heavy for its contribution to dominate the X-ray intensities. It is then an easy matter to find its position in the unit cell by computing a Patterson synthesis, which will clearly show heavy atom-heavy atom vectors. One proceeds to calculate the diffraction pattern, both amplitude and phase, which such an atom would produce if it were the *only* atom in the unit cell. The result will resemble the observed pattern, but naturally will not be identical to it, since the contribution of the rest of the molecule has been omitted. Now the observed diffraction pattern gives us the correct *amplitudes* for the whole structure, but not the phases. One employs, therefore, as the next best thing, the *calculated* phases—based on the heavy atom alone—together with the *observed* amplitudes, to calculate a Fourier or electron density synthesis. This will show the heavy atom and in addition a "ghost" of the rest of the molecule

(more often, two ghosts superposed—one left-handed, the other right-handed). With luck this rather confused picture enables one to guess the positions of the atoms making up the molecule. A new set of phases can then be calculated, based on them as well as on the heavy atom. From then on the process is one of refinement, of making successive small shifts in the atomic positions to improve the overall agreement of calculated and observed amplitudes.

The major difficulty in applying this method to proteins would be that no single atom is heavy enough to control the majority of the phases because of the large size and scattering power of even the smallest protein molecule. Nevertheless, as we shall see, it may prove possible to circumvent this limitation to some extent by using a large number of heavy atoms attached simultaneously to the protein molecule.

b. The Method of Isomorphous Replacement. This method is often loosely referred to as the Heavy Atom method, but should be clearly distinguished from the foregoing; it is better described as the method of isomorphous replacement. It is not too much to say that the subject of protein crystallography has been transformed by the pioneer application of this method to hemoglobin by Perutz and his collaborators. It requires *two* practically identical unit cells, one containing a heavy atom (preferably one per asymmetric unit, but see p. 194 below) such as mercury, and the other having all the atoms in the same places as in the first, with the exception of one or two light atoms (often a water molecule) in the place where the heavy atom was before. This requirement is not easy to meet in proteins. One wants substitution at a unique site on the surface of molecules which in general do not contain many unique sites—it is only in special cases, such as proteins with prosthetic groups or sulfhydryl groups, that they are obviously present. On the other hand protein crystals present advantages from another point of view, because of the considerable volume of the unit cell occupied by solvent, providing many places where extra molecules can be added without disturbing the arrangement of the protein.

As in the Heavy Atom Method, the presence of our mercury atom (for example) will change the intensities of the reflections in the diffraction pattern. In considering the nature of the changes produced, and how they may be used to investigate the structure of the protein, we shall in the first instance restrict ourselves as follows. It has been explained (see p. 136) that any particular reflection has a certain amplitude and phase. It can therefore be regarded as a vector; moreover, the contributions to it of all the atoms in the unit cell are themselves vectors, which must, of course, be combined vectorially. However, in certain planes of the reciprocal lattice (depending on the symmetry of the crystal), corresponding

to certain special projections of the structure, these vectors are all either parallel or antiparallel, and can therefore be added arithmetically. We may speak of such reflections as being *real*, i.e. having no imaginary vectorial component; as having phase angles of 0 or π; or as being positive or negative. The corresponding projection is known as a *real* projection. Our first discussion will confine itself to these reflections.

Consider, then, a certain (*real*) reflection, and suppose that its intensity (on some arbitrary scale) is 100 for the case where there is protein only, and 64 for the case where there is the same arrangement of protein and in addition one mercury atom per asymmetric unit. The amplitudes will be the square roots of these numbers, that is 10 and 8 respectively; but since we do not know whether the phases of the reflections are 0 or π we must write these as ± 10 and ± 8. The mercury contribution must be the difference of these two numbers, that is to say either ± 18 or ± 2. For most reflections the contribution of the mercury is smaller than that of the protein, so the correct value will be the smaller one of the pair, that is ± 2 in our example. We are still left with an ambiguity of sign, however: that is to say we have, to correspond with

	protein	+	heavy atom	=	(protein plus heavy atom),
either	(+10)	+	(−2)	=	(+8)
or	(−10)	+	(+2)	=	(−8)

Which of these is correct we cannot yet tell. What we do know, however, is that the amplitude due to the heavy atom is ± 2; and thus if we could have a unit cell with all the protein subtracted, empty except for the heavy atom alone, the intensity of the reflection we are considering in its diffraction pattern would be $(\pm 2)^2 = 4$. We can thus calculate, without making assumptions, the intensities in the diffraction pattern due to the heavy atom alone—the so-called *difference intensities*. To find the position of the heavy atom we carry out a Patterson synthesis, known as a *difference Patterson synthesis* or $(\Delta F)^2$ synthesis. This shows us the vectors between the heavy atoms in each of the asymmetric units of the cell, all the vectors involving atoms other than the heavy atom having been canceled out; and from it one can simply obtain the position of the heavy atom relative to the symmetry elements of the cell. Examples are given in Figs. 16 and 18.

Having found the heavy atom we can calculate its contribution to any particular reflection. Let us suppose that it comes out to +1.7 for the reflection we have been considering. Allowing for experimental error this agrees with the second of our two alternatives; it follows that the *protein* reflection must be −10 and not +10. That is to say, *we have determined the phase of this particular reflection.* If we can do this success-

fully for all the reflections in the reciprocal lattice plane we have been studying, the way is open to calculate an electron density map of that particular projection of the structure.

So much for the real projections. For those regions of the reciprocal lattice where the reflections are complex (that is, of general phase) and which comprise its major part, we follow an analogous but more complicated procedure, whose results are less definite.

We cannot do as we did before, that is subtract the two amplitudes, because the corresponding vectors are not parallel. It turns out that to find the heavy atom we have to calculate another variety of Difference Patterson, called a (ΔI) synthesis, in which the terms are simply the difference in *intensity* between (protein + heavy atom) and protein alone. In algebraic terms we use

$$\Delta I = I_{P+H} - I_P$$

where P = protein; P + H = protein plus heavy atom, whereas before we used

$$(\Delta F)^2 = (\sqrt{I_{P+H}} - \sqrt{I_P})^2$$

The ΔI type of Difference Patterson synthesis gives us as before the vectors between the heavy atoms, from which we can calculate their co-ordinates in the unit cell; but superposed on them are all the vectors between heavy atoms and *every other atom in the unit cell*, i.e. atoms of protein or liquid (although our procedure does remove those between protein and protein). The diagram thus has a confused background and it may be difficult to locate the heavy atom-heavy atom vectors.

Even when we have successfully done this there are still further difficulties. Without going into details we may say that:

1. With one single isomorphous replacement we cannot hope to find the phases of all the reflections directly. What we get is two values for the phase angle of each reflection, and to remove this ambiguity a second isomorphous replacement in a different place in the unit cell is necessary. To be on the safe side it would be better to have at least three separate isomorphous replacements.

2. The accuracy is less than in the case of real reflections, since we have to assign a quantitative value to the phase angle—a value which will be in error—whereas in the real case all we have to do is to decide between two alternatives, plus or minus.

In practice, therefore, working with reflections of complex phase is quite a different proposition from working with real ones. It is less accurate and more troublesome; besides, the actual number of general reflections of general phase is far greater. On the other hand, experience with the few proteins where electron density maps have been produced indicates, as we shall see, that a two-dimensional projection of the unit cell is very little use even when it is known to be correct—the thickness of protein and solution through which the projection must be made (never less than 30 A., representing

15–20 atoms) is so great that all the features of interest are obscured and the result is an uninterpretable confusion. To make real progress the third dimension *must* be broken into, even though the labor involved is at best formidable.

It should be added that one of the inescapable difficulties of protein crystallography is that it is impossible for *all* the reflections from a protein crystal to be real. This could only happen if the mirror-image protein molecule (made up of *dextro* residues, and related to the real molecule as a right-hand glove is related to a left-hand glove) were also present. So three dimensions mean solving the general phase problem. On the other hand some space groups are more favorable than others for studying projections; thus monoclinic unit cells have one real projection, whereas orthorhombic ones have three, mutually perpendicular.

To illustrate the use of the method we shall now describe some of the results obtained by it so far.

c. Isomorphous Replacement and the Structure of Hemoglobin. This was the first application of the method. Perutz and his collaborators (Green *et al.*, 1954) made use of the fact that hemoglobin contains free sulfhydryl groups by causing it to react with *p*-chloromercuribenzoate (PCMB), a standard reagent for SH groups. In this way they obtained a hemoglobin molecule with two mercury atoms attached to its surface at specific and definite sites. After crystallization the dimensions of the unit cell were found to be quite unchanged, but the X-ray photographs showed unmistakable changes in the intensities of the reflections. The reflections corresponding to a projection along the *b* axis, which in this space group are all real, were measured carefully for both normal and mercury-substituted hemoglobin. The two sets of intensities were adjusted to the same scale by a statistical method and the difference between their square roots (i.e. their amplitudes) gives $|\Delta F|$, the change in amplitude produced by the mercury atom. A difference Patterson projection was computed using values of $(\Delta F)^2$; it is shown in Fig. 16. It will be noticed that apart from the peak at the origin (always present in Patterson syntheses, and merely representing the fact that every atom in the unit cell is at *zero distance from itself*), there is one other peak much larger than all the rest. It has the coordinates (14.8, 31.6). It follows that in the unit cell the x and z coordinates of the heavy atoms are (+7.4, +15.8) and (−7.4, −15.8) relative to an origin at the dyad axis of symmetry by which they are related.

In fact, however, there is more than one solution since this unit cell has two different dyad axes and two different screw dyad axes (which appear like dyads in projection). The electron density map derived from the reflections would look just the same whichever solution were adopted; it is only the symmetry axes which would be wrongly labeled—dyads instead of screw dyads and vice versa—if the wrong solution were selected. Fortunately in most space groups the ambiguity does not arise.

As it happens Perutz was able to decide which solution was correct by making use of the extensive experimental results on the shrinkage and expansion of the crystal. Horse hemoglobin crystals are unusual in undergoing a particularly simple type of shrinkage. The molecules lie in sheets and during shrinkage each sheet remains quite unchanged in itself, but moves relative to its neighbors, in a direction always perpendicular to the *b* axis. Using this fact it can be shown (Bragg and Perutz, 1952c;

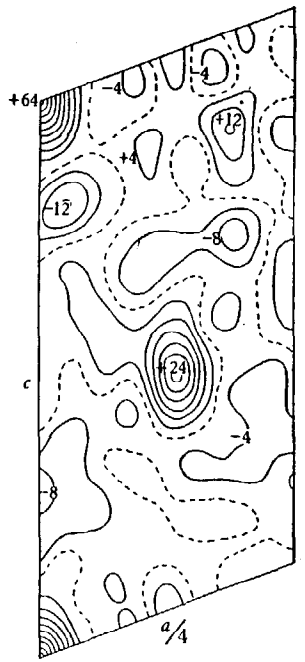

FIG. 16. A Difference Patterson, to show how the position of a heavy atom is discovered using the isomorphous replacement method. The origin is shown twice, at the top and bottom left-hand corners. The peak representing the end of the vector *between* heavy atoms is near the middle of the map (labeled + 24). The other (smaller) peaks and hollows are spurious background due to errors of measurement, etc. (Horse hemoglobin: differences due to PCMB. Green *et al.*, 1954.)

Perutz, 1954) that certain groups of reflections must have their signs linked together. Moreover, from the study of the many different crystal forms of hemoglobin, and of the changes due to variations in the salt concentration in the mother liquor, Bragg and Perutz (1952a) had been able to deduce the approximate shape of the molecule and its position in the cell, and thus to decide which particular dyad axis relates the two halves of the molecule to one another.

The position of the mercury atom having been found, the next step was to calculate its contribution to each reflection. The sign of the protein reflections could then be deduced by inspection in the way we have already

discussed: that is to say, if the X-ray picture showed that the mercury had increased the intensity of a reflection, then it was given the same sign as the calculated mercury contribution; if the intensity had been decreased, then the signs were made different. In some cases the mercury contribution happened to be so small that no decision could be made, but in the great majority of reflections the signs could be definitely allocated.

The earlier studies, predicting that certain reflections would be of like or unlike sign, were confirmed wherever they made definite predictions. Some of the weaker predictions were not confirmed, however, and it is because of this, and because this method of linking signs is only possible in very special cases, that we have not described it in detail. It was nevertheless a technical *tour de force* at the time.

Recent work, which will be mentioned shortly, has increased the number of definite sign determinations, and confirmed those allocated earlier, with the result that 88 reflections out of a total of 94 whose spacings exceed 6 A can now be taken as certainly established. From these data an electron density map has been computed, showing the contents of the cell projected parallel to its *b* axis. By combining the results of the various shrinkage stages Perutz was able to calculate his electron density projection corresponding to an imaginary superexpanded stage in which the layers of molecules have been, as it were, floated apart, so that there is open water between them. (Note that this procedure is only possible in very special cases, as we have indicated above. In general one cannot separate out the molecules from their overlapping neighbors.) The result, showing the projection of a single layer of hemoglobin molecules, is reproduced in Fig. 17. This has been drawn in such a way that the zero contour represents the electron density of water: protein is, on the average, more dense than this.

Looking at Fig. 17 one experiences two feelings: admiration for the very considerable technical achievement which it represents, and disappointment that the result appears to give us so little information. Its obscurity is due mainly to the very great thickness of the projection—the unit cell is 63 A thick in this view—and partly to the rather low resolution. Nevertheless there are some interesting features. For example, the outline can be fitted roughly to the ellipsoidal shape deduced by earlier methods (see p. 178), but is markedly more irregular. Again, the molecule appears to have a waist, or rather a dimple, close to the dyad axis; presumably this is related to the fact that it consists of two identical halves. None of the other features suggest anything in particular, though the "hole" marked *w* is surprisingly, though not impossibly, deep. The features which a biochemist might first search for—the iron atoms and the heme groups—would not in any case be expected to show up in projection at this resolution.

The one feature which can be identified with certainty is the position (in projection) of the heavy atom, although even this could not be picked out if we did not have two views, one with and one without it (the figure shows the latter). Since we know that the mercury atom is attached to a sulfhydryl group, it is possible to draw some conclusions about the position of sulfhydryl groups in the hemoglobin molecule by a correlation of chemi-

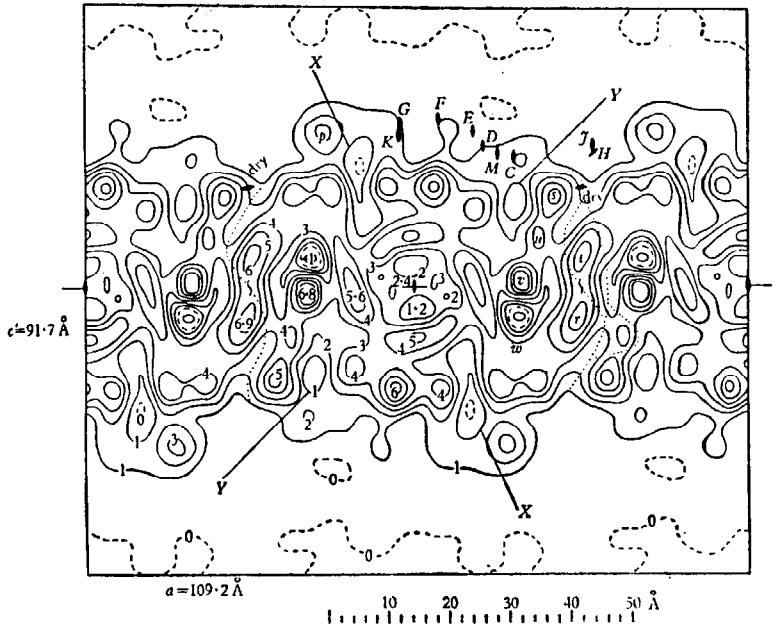

FIG. 17. A Fourier projection of a row of hemoglobin molecules suspended in salt-free water. The contours are contours of (projected) electron density, the zero contour corresponding to the density found where the whole depth of the unit cell is filled with water. The dyad axis between the two halves of the molecule is in the center of the figure. (Bragg and Perutz, 1954.)

cal and X-ray studies. We shall enlarge on this topic since, one hopes, it provides the pattern for much future work.

It is convenient to summarize the chemical work first (Green *et al.*, 1954; Ingram, 1955), fixing our attention for the moment on native horse hemoglobin.

Ingram used the technique of amperometric titration with silver nitrate: thus the protein would first be made to react with a known quantity of PCMB, and the SH groups remaining unblocked would be titrated with $AgNO_3$. The results showed that all the available sulfhydryl groups of a single hemoglobin molecule (molecular weight 67,000) could be saturated either by 4 molecules of $AgNO_3$, or by 2 of PCMB, or by 2 of $HgCl_2$ (the experiments were carried out under conditions such that combination of the reagents was probably with SH groups, though this is not certain). If the molecule was first saturated with $AgNO_3$ or with PCMB subsequent reaction with $HgCl_2$ displaced the first substituents. If on the other hand *one* mole of either PCMB or $HgCl_2$ were added first the protein would subsequently take up only *two* moles of $AgNO_3$, the reagent first added not being displaced in this case. The surprising thing is that stoichiometrically PCMB and $HgCl_2$ are equivalent, although one would expect the former to be univalent relative to SH, and the latter divalent.

The results suggest that native horse hemoglobin contains *four* available sulfhydryl groups, *arranged in two close pairs*. One molecule of $HgCl_2$ or PCMB will saturate *both* the SH groups of one pair, the former by combining directly with both of them, the latter by combining with one and inactivating the other by steric hindrance. On the other hand a silver atom, being much smaller, saturates only one SH of a pair, and *two* are required to inactivate the pair altogether. One would expect, therefore, that there would be only two regions on the hemoglobin molecule where mercury or silver would go, each corresponding to one of the pairs of SH groups. The X-ray results confirm this, at least as far as the x and z coordinates are concerned. Difference Fourier projections have been prepared, showing the positions of the heavy atoms, for the complexes of hemoglobin with (a) 2 moles of PCMB, (b) 2 moles of $HgCl_2$, (c) 2 moles of $AgNO_3$, and (d) 4 moles of $AgNO_3$. The resulting projections all show heavy atoms combined at approximately the same positions in the cell. It is especially significant that when *four* silver atoms are combined the unit cell contains only *two* peaks, showing that they are present in two pairs, each pair being so closely spaced that at 6 A. resolution the two silver atoms composing it cannot be seen as separate peaks.

This is not the whole story, however. Analytical data for horse hemoglobin (Tristram, 1953) show that it actually contains 6 sulfur atoms in the form of cystine or cysteine. Other experiments by Ingram have shown that, *when denatured*, hemoglobin can react with *six* moles of $AgNO_3$, in contrast to the native protein which, as we have said, reacts with four. This result, together with others on blocking by $HgCl_2$ and by PCMB, suggest that the SH groups actually occur in two groups of *three*, but that one member of each group is unavailable in the native protein. Ox, sheep, and human hemoglobin have also been studied with similar (but not identical) results; Ingram's paper (1955) should be consulted for details.

Careful study of the X-ray data shows that although in all the derivatives we have mentioned the heavy atom is attached to the same part of the molecule, there are in fact minor differences in position, amounting to a few Angstrom units. The reason for these small differences is unknown; they may be due perhaps to rotations about the bonds of the cysteine side chain. They were in fact of great assistance in sign determination; the signs of some of the reflections could not be decided from the PCMB derivative alone because the mercury happened to be in such a position

that its contribution to their amplitudes was almost zero, while the silver atoms were displaced sufficiently to ensure that in such a case their contributions would be quite substantial.

At the time of writing nothing further has been published on hemoglobin. Sign determination has, however, been extended to a resolution of 3 A. and a Fourier projection with this resolution has been calculated; as might have been anticipated it does not reveal any additional features of the structure which can be interpreted, at present at least (Perutz, personal communication).

Perutz (1956) has worked out the theory of a method for determining the difference in y coordinates between two heavy atoms used in two separate isomorphous replacements (the minimum requirement for three-dimensional work—see p. 182); in a monoclinic cell there is no simple way of establishing this difference, since there are no symmetry elements perpendicular to y which can act as reference points.

Work has also been in progress on ox hemoglobin; Green and North (personal communication) have obtained several isomorphous replacements, again using the sulfhydryl groups, and have determined a substantial proportion of the signs of the a and c projections of the (orthorhombic) unit cell. They have also derived a tentative Fourier projection along x, a view of the molecule already known to have interesting features (Crick, 1953a, 1956). In both species of hemoglobin the most pressing problem is to obtain further isomorphous replacements at radically different places on the surface of the molecule—in this endeavor some success has been achieved by the use of two reagents developed for the work on myoglobin (see p. 191), namely mercuriiodide and aurichloride, but it cannot be said that the problem is yet entirely solved. Its solution would open the way for determining the exact shape of the molecule in a fairly short time, and in the long run for a full three-dimensional analysis.

d. Isomorphous Replacement and the Structure of Myoglobin. From several points of view myoglobin is an attractive object for study by the protein crystallographer. It has a small molecular weight (17,000) and it can readily be crystallized in at least a dozen different space groups (Kendrew *et al.*, 1954). It contains a prosthetic group of known chemical structure and defined physiological role, and it is analogous in function and so perhaps in structure to another protein which is being intensively studied by X-ray methods, namely hemoglobin. On the other hand it is more intractable from the point of view of isomorphous replacement, because no myoglobin is known to contain free sulfhydryl groups, whereas these groups are to be found in all hemoglobins so far examined. The other obvious unique site in the molecule is the heme group itself, and various attempts were made (Kendrew, Bodo, Dintzis, and Ingram, unpublished data) to prepare ligands for the heme group which contained heavy atoms such as mercury and iodine. Imidazoles, nitroso compounds and isocyanides were all investigated, but for various reasons none of them was wholly successful—generally the reason was that myoglobin has so high an affinity for gaseous oxygen (much higher than that of hemoglobin), with the result that unless the most strictly anoxic conditions were maintained the heme group ligand was promptly replaced by an oxygen molecule.

In the end success was achieved by an entirely different approach, namely to crystallize myoglobin in the presence of various inorganic ions containing heavy elements. Naturally ions were chosen which on general chemical grounds might be expected to have some affinity for one or more of the types of side chain present in proteins. But of course a protein nearly always contains more than one of any given side chain; the hope was that in some cases steric or other factors might induce the ion to be attached preferentially or even specifically at a single site. In effect this hope was realized in a number of instances. The criteria for success were crystallographic rather than chemical: that is to say, an X-ray picture showing changed intensities was the evidence that combination had taken place; and a difference Patterson, computed from the intensity changes in the same way as we have described above, indicated that combination had been specifically at a single site if it was found to contain only one peak per asymmetric unit. We may take as an example the first ion that was successfully attached in this way—namely mercuriiodide HgI_4^{--}. This was investigated because it was known to form complexes with thio ethers; myoglobin contains two residues of methionine whose side chain is $—CH_2—CH_2—S—CH_3$. Myoglobin crystals prepared in presence of potassium mercuriiodide gave an X-ray pattern substantially different from normal, and the Difference Patterson projection calculated from it is shown in Fig. 18. (We shall be dealing throughout with myoglobin derived from sperm whale and crystallized from ammonium sulfate. It is known as Type A and is monoclinic, with two molecules in the unit cell; the space group is $P2_1$, which means to say that the only symmetry elements present in the unit cell are screw dyad axes parallel to b.) Figure 18 contains only one peak per asymmetric unit (i.e. per half cell) apart from the origin peak, and it may therefore be taken that combination has occurred at one site per molecule. This is not an expected result since sperm whale myoglobin contains two methionine side chains, not one; it must be supposed that one of them only is sterically available for combination—if indeed the methionine side chain is the site of attachment. The difference Patterson is computed from reflections of spacing greater than 4 A.; the mercuriiodide group is therefore not resolved into its component atoms. A later projection with all terms out to 2 A. (not illustrated here) shows the group partially resolved.

From the known position of the mercuriiodide group many of the signs

of the reflections could be deduced in exactly the same way as was done with hemoglobin. To be certain of all of them, however, more than one isomorphous replacement was required. In fact several other methods have been discovered; in most of them the chemistry involved is even more obscure than it is in the case of mercuriiodide. Figure 19 shows the positions of the different replacements in the unit cell. In each case they were located by means of a difference Patterson or difference Fourier projection. In some cases the method of achieving isomorphous replacement was

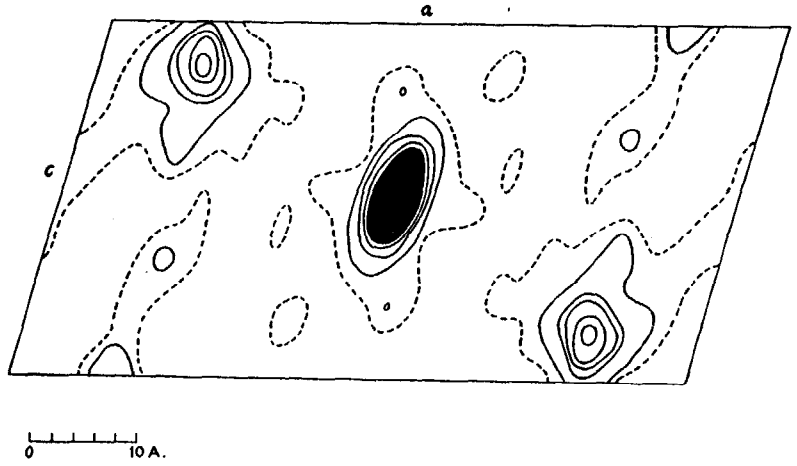

FIG. 18. Another example of a Difference Patterson. This time the origin is in the middle of the figure. The vector between heavy atoms produces the peak near the top left-hand corner (and also, by symmetry, near the bottom right-hand corner). It can be seen that the distance between either of these peaks and the origin is about 23 A, showing that in projection each heavy atom is 11½ A from a twofold (screw) axis, and thus 23 A from the other heavy atom. (Sperm whale myoglobin, type A; heavy atom due to K_2HgI_4. Bodo, Dintzis and Kendrew, unpublished data.)

extremely simple—for example, the silver atom indicated in the figure was introduced by the simple expedient of crystallizing myoglobin in the presence of 1 mole of silver nitrate. Each of the replacements provides an independent check of some 80% of the signs; and this means that all reflections except a few weak ones are at least triple checked. Out to spacings of 6 A. there are no discrepancies: beyond this the analysis has not been carried in detail.

Two of the ligands shown in the figure are specific reagents for the heme group, namely *p*-iodonitrosobenzene and a compound made by combining the mercury atom of *p*-chloromercuribenzene sulfonate with the sulfhydryl group of *p*-mercaptophenylisocyanide. Neither of these gave interpretable

Difference Patterson projections, any more than the other specific heme group reagents which were tried; but once all the signs of the reflections had been established by other means it was possible, making use of them, to locate the heavy atoms in these combinations by Fourier methods. As the figure shows, they are in almost identical sites. Since the two compounds react specifically with the iron atom of the heme group, we may conclude that the latter is situated about 6 A. from this site.

The next contour map (fig. 20) is a Fourier projection of the protein with a resolution of 6 A. A Fourier projection with resolution 4 A. has

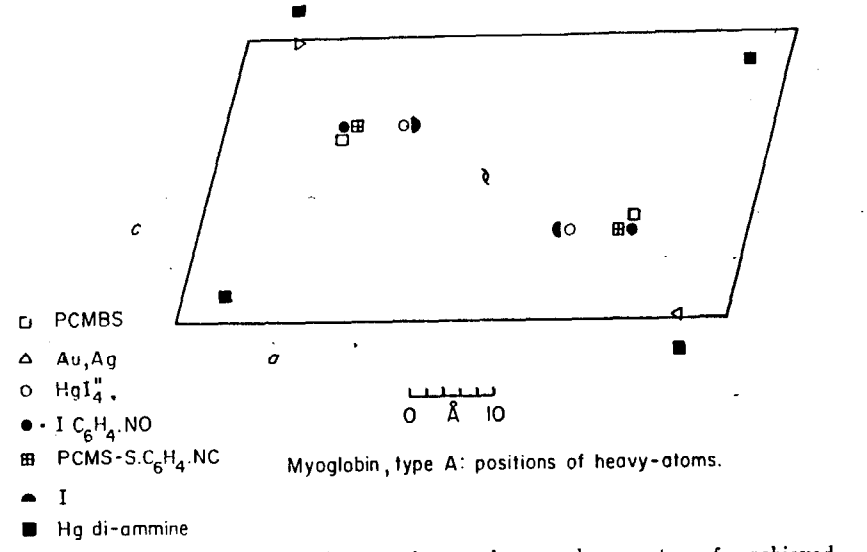

FIG. 19. To show the seven different isomorphous replacements so far achieved in myoglobin, type A. The symbols indicate the positions of the different heavy atoms in the unit cell. (Bodo, Dintzis and Kendrew, unpublished data.)

also been computed, but is almost identical, indicating that the confusion of peaks and hollows in the projection which makes it virtually uninterpretable is a consequence not so much of inadequate resolution but of the great depth of protein (31 A. in myoglobin) through which the projection is necessarily made. We have here one more example of the fact that two-dimensional projections of protein unit cells are not at all informative, and once again it is clear that a three-dimensional approach, with all that is involved of tedious computation and greater error, is quite essential.

Besides the work on Type A myoglobin crystals, attempts have been made to achieve isomorphous replacement in several other crystalline forms of the protein. The most successful of these is seal myoglobin

(Type C; Scouloudi and Kendrew, unpublished data), a monoclinic form of which a preliminary Fourier projection, based on signs deduced from isomorphous replacement with mercuriiodide, has been computed. Not all the ions used with Type A are equally successful with Type C, however;

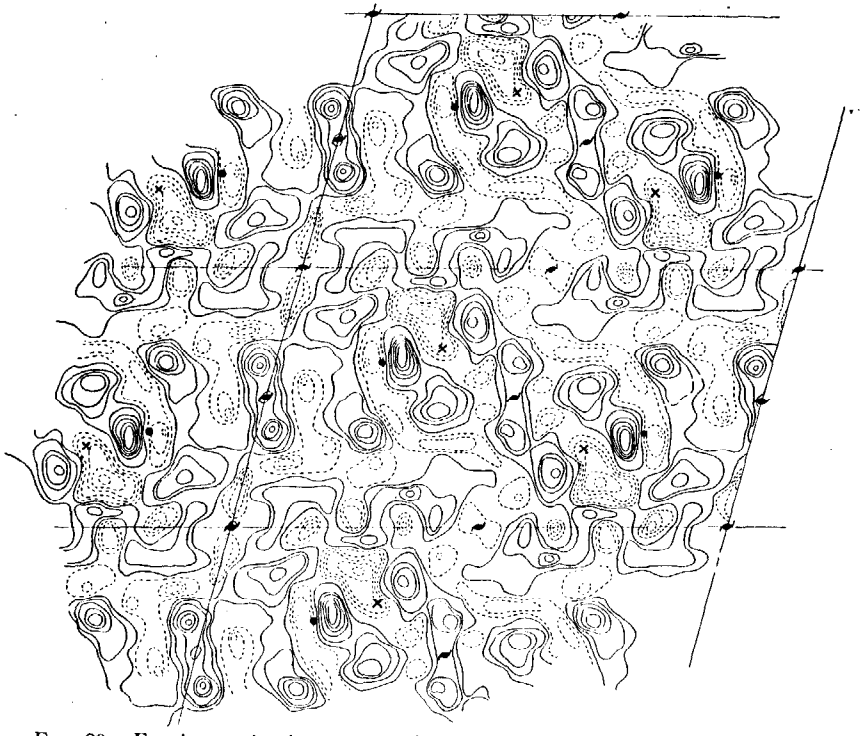

FIG. 20. Fourier projection of myoglobin, type A. This shows the projected electron density of the monoclinic unit cell (which contains two molecules), looking down the *b* axis, i.e. through 31 A of protein and mother liquor. Resolution 6 A. The position of two heavy groups are shown: •HgI_4^{--}; × Hg in *p*-chloro-mercuribenzene sulfonate. The two molecules overlap and it is not easily possible to tell where one begins and the other ends. (Bodo, Dintzis and Kendrew unpublished data.)

and this suggests that in some cases at least the "combination" between protein and ion may be interstitial rather than strictly chemical; different methods of packing virtually the same molecules in crystalline array would lead to the formation of different "cozy corners," apt for the harboring of different small ions.

e. Isomorphous Replacement and the Structure of Ribonuclease. Ribonuclease is another favorable protein for X-ray studies because of its

low molecular weight (about 13,500), because of its ready availability, and because a great deal is already known about its amino acid sequence (for a summary of the present position see Anfinsen and Redfield, 1956); indeed it is likely that the whole sequence will shortly be established. There are several different types of crystal (King *et al.*, 1956), but most work has been carried out on a monoclinic form (space group $P2_1$, two molecules per cell). Earlier tentative interpretations of the Patterson projections of this form (Carlisle and Scouloudi, 1951; Carlisle *et al.*, 1953) have not been widely accepted by other workers in the field. The *three*-dimensional Patterson syntheses (Carlisle and co-workers, unpublished data; Magdoff *et al.*, 1956) in fact show less "regularity" than those of hemoglobin, myoglobin, and insulin. It is thus not very likely that the structure will be solved without the help of either the isomorphous replacement or the heavy atom method, even when the amino acid sequence is known.

No completely successful isomorphous replacement has yet been reported in print, but Dr. David Harker and his co-workers (personal communication) have succeeded in introducing into the crystal a uranium complex of the dye alizarine cyanone RC, in such quantity that the unit cell contains four dye molecules and eight uranium atoms. A preliminary Fourier synthesis of the *b* projection of the unit cell has been obtained (Harker, 1956). It is understood that further isomorphous replacements are being attempted both on the monoclinic crystal form (Carlisle and co-workers) and on an orthorhombic form (Harker and co-workers).

f. Requirements for Isomorphous Replacement. Having described some applications of isomorphous replacement to proteins we shall now state the basic requirements of the method. The first of these is the obvious one that the heavy atom should be heavy: but, we may ask, how heavy? The parameter measuring the effective "heaviness" is approximately given by

$$M_h \sqrt{\frac{n}{M_p}}$$

where n = number of heavy atoms per protein molecule
M_h = atomic weight of heavy atom
M_p = molecular weight of protein.

Two points must be made about M_h. The first is that in practice the heavy atom replaces a light molecule, probably of solvent, so that something should be subtracted to allow for this. The second is that very often the heavy atom does not fully saturate its site; if only 80% of the sites are occupied then the heavy atom will appear to the X-rays to be less

heavy, in proportion. Our formula is therefore better written

$$k(M_h - 25)\sqrt{\frac{n}{M_p}}$$

where k = occupancy by the heavy atom (= 1 if all sites occupied). (If the n heavy atoms are very close together compared with the spacing of the reflection being considered, it can be shown that $\sqrt{n}$ should be replaced by n in the above expression).

As an example we may take the case of mercury added to horse hemoglobin, where n was 2 per molecule of molecular weight 68,000 and k was believed to be 0.8. Our parameter is then

$$0.8(200 - 25)\sqrt{\frac{2}{68000}} = 0.76$$

If we add one atom of mercury to one molecule of myoglobin, with 100% occupancy, the parameter works out to be 1.35.

It is still not clear what is the lowest useful value of the parameter, but there is no doubt that a value 1 is satisfactory for real structure factors, while one of 0.5 would be marginal; for complex structure factors higher values are desirable—1.5 or 2.0 or even larger. Thus in practice a single bromine atom, for example, is near the lower limit even for a small protein. For complex phases it is an advantage to have more than one heavy atom at a time, but multiplicity of heavy atoms raises new problems which we shall now consider.

To specify the most desirable number of heavy atoms is not a simple problem. In the first instance it is probably best to have only one heavy atom per asymmetric unit, since it can be located with great ease. Nevertheless it may be possible to find two, or even more, especially if some of them are linked together chemically so that their distance apart is known. It is difficult to be dogmatic, but four or more would certainly be difficult to locate, though if the atoms were really heavy, e.g. mercury or uranium, the difficulties would be reduced somewhat.

However, once the signs of the protein reflections in a given projection have been established by an initial simple isomorphous replacement, more complex replacements can be dealt with relatively simply, since one can then work in real space by computing difference Fourier projections, showing directly where the heavy atoms are, rather than difference Patterson projections which become very complicated if n is much above one (the number of peaks in a Patterson projection is the *square* of the number of atoms in real space). Under these circumstances there should be no difficulty in locating quite a number of heavy atoms, say 10 or even 20, assuming that isomorphism is maintained with so large a number of foreign atoms in the unit cell. For multiple replacements of this kind the greatest difficulty would be to find their y coordinates in a monoclinic cell. So large a number of heavy atoms, if accurately located, would be of the greatest value for three-dimensional work, and it might even become possible to use the Heavy Atom Method proper (see p. 179).

We may summarize the requirements for isomorphous replacement as an aid to structure determinations:

a. A substantial degree of saturation is desirable at all sites which are occupied at all. It makes every kind of application difficult if, in addition to some sites being fully occupied, there are others slightly occupied—for the latter cannot readily be distinguished from experimental error.

b. It is a great advantage to have several *different* isomorphous replacements on the same protein; that is to say, replacements on different parts of the molecule.

The minimum useful separation between two sites, if they are to be considered as "different," depends on the spacing of the X-ray reflections being studied—for long spacings they must be further apart than they need be for short spacings; but a separation of as little as 5 A. can give much useful information, though not about the inmost reflections. It follows that there is scope for adding different molecules, containing heavy atoms in different places, to the same point of attachment on the protein.

c. At least one of the isomorphous replacements should give a large value to the parameter discussed above. That is, *the atoms should be really heavy.*

If these criteria can be satisfied there seems at present no insurmountable obstacle in principle against proceeding toward a complete structure determination of a crystalline protein. At the very best, however, this will be a very long and tedious business. It may be asked whether in the short run there are any quick returns which can be expected from an investment in heavy atoms, in advance of the long-term dividends which protein crystallographers patiently hope to receive. There appear to be two ways in which quick answers, of direct use to biochemists, can be obtained by means of isomorphous replacement; the determination of the shape of the molecule, and the localization of interesting groups on its surface.

We think that methods involving isomorphous replacement are likely to become the standard ones for determining the shapes of all except perhaps the very large protein molecules. Since only the low order X-ray reflections are involved (high resolution is not needed) strict isomorphism is not necessary, and the amount of experimental data which has to be collected is not very large; on the other hand at least two and probably three separate isomorphous replacements are needed. If these are available, and especially if the electron density of the solvent can be varied—as is usually the case—it should be possible to obtain a good low resolution three-dimensional picture of the molecule.

We do not attach much importance to the objection that proteins may change their shape while passing from solution to the crystal, with the implication that results obtained by crystallographers are not relevant to the interests of biochemists; as we have explained, the environment of the protein in a crystal is very similar to its environment in solution. We do agree, however, that some changes may sometimes occur (see Yang and Doty, 1957); but even so, firm information about a slightly altered shape is very much more useful than dubious information about the actual shape under physiological conditions.

A more important difficulty is that where two neighboring molecules are touching it may not be easy to discover from a low resolution picture where one ends and the other begins. However, such ambiguities can probably be resolved by the comparison of different shrinkage stages or different crystalline forms of the same protein.

The second type of quick answer which one may hope to obtain by the help of heavy atoms is the position of side chains or active centers which can be "marked" by specific attachment of groups containg heavy atoms. This procedure, while at first sight very attractive, may not always prove to be of real value, since the position of the heavy atoms is found in the first instance only in relation to the symmetry elements of the crystal. That it can sometimes give useful information may be gathered from one example that we have already considered—the location of sulfhydryl groups in hemoglobin. In any case, as isomorphous replacements on the same protein multiply, the interrelations of different groups begin to emerge and acquire significance; and in addition the shape of the molecule and the positions of the heavy atoms within it should become available. The kind of problem which might easily be solved, for example, would be the establishment of the relative positions of the active center of an enzyme and of a sulfhydryl group in the molecule: the former might be "marked" by means of a specific inhibitor containg a heavy atom, the latter by methylmercury or PCMS. There may also be a point in adding heavy atoms to localize particular side chains or active centers, even if the resultant complex does not crystallize isomorphously with the normal protein. Thus if the tyrosine residues in a protein were to be iodinated, the result might be that the cell dimensions of the crystal changed substantially, or even that the space group was altered; but it might nevertheless pay to work with the iodinated protein in other isomorphous replacement studies, since eventually it would be easy to locate the iodine atoms, and these would immediately reveal the position of the tyrosine residues. If the amino acid sequence of the protein were known this would be of considerable help in building trial models of the structure.

There are thus several ways in which heavy atoms may be added to a protein so as to be useful for X-ray diffraction studies. Each case has to be examined individually on its merits, especially since there is as yet no means of predicting whether a particular addend will alter the size of the unit cell. It seems to us that these requirements present a challenge to the protein chemist. To attach heavy atoms at relatively few sites and in high yield, without denaturing the protein, is not an easy task; nevertheless since the information which can be obtained if success is achieved is potentially very great the expenditure of considerable effort is justified. We hope that our short outline of the problem will stimulate other workers to devise new methods to this end.

5. The Chain Configuration in Globular Proteins

As we saw in a previous section, there is good evidence that the synthetic polypeptides, in their folded form, assume the α-helix configuration of Pauling and Corey; further, it is very probable that certain fibrous proteins contain the α-helix; among these may be mentioned α-keratin, tropomyosin, myosin, fibrin, and intact bacterial flagella. The question which we shall now discuss is how strong are the many claims that the α-helix is also the main structural feature of the globular proteins. These claims are controversial, and in our view the case is not yet proved. We shall summarize the evidence briefly.

Similarities between the X-ray pattern of α-keratin and the Patterson synthesis of hemoglobin were pointed out by Perutz (1949): parts of the three-dimensional vector structure contain rods spaced 10 A. apart with maxima at 5 A. intervals along their length, and the suggestion was that these rods corresponded to polypeptide chains in real space, whose dimensions would correspond to those in α-keratin. Similar rods have been found in myoglobin (Kendrew, 1950; Kendrew and P. J. Pauling, 1956). When the α-helix was discovered Pauling and Corey (1951b) worked out the radial average vector density of Perutz's Patterson synthesis of hemoglobin, and showed that it was not unlike what one might expect from an assembly of α-helices. Their comparison was in effect a comparison between theoretical and experimental versions of the powder pattern of the protein—that is to say, the pattern which would be obtained by irradiating a random assembly of hemoglobin molecules. This comparison has been made directly by Arndt and Riley (1955), who measured the intensity of X-ray scattering as a function of angle for a large number of amorphous protein specimens, and concluded that the α-helix was an important constituent of most of them, hemoglobin and myoglobin included.

In order to prove, by this method, that the α-helix is a predominant structural feature of a protein, two criteria would have to be satisfied. First, the common features in the powder patterns of those proteins for which the hypothesis is made should correspond exactly, or nearly exactly, with those known to be produced by the α-helix (as shown either by calculation or by comparison with a variety of synthetic polypeptides known to be in the α-form). Second, no other possible chain configuration should

give these features. The matter is a highly technical one, but in our view the evidence presented is unsatisfactory on both these counts. Besides, Arndt and Riley conclude that their observed curves are best fitted by the *left-handed* α-helix; very recently, as we have already pointed out (p. 157), evidence has been forthcoming that the stable form of an α-helix of L-residues is *right-handed*.

There have been various attempts (mostly unpublished) to discover a 1.5 A. reflection in the diffraction pattern of globular proteins, this being a diagnostic feature of the X-ray pictures of fibers containing the α-helix (see p. 155). Apart from a very diffuse "spot" in the diffraction pattern of horse hemoglobin (Perutz, 1951), no such reflection appears to have been found, although searches have been made in insulin (Low, 1955), ribonuclease (Bernal and Carlisle, personal communication), and myoglobin (Kendrew and P. J. Pauling, unpublished data).

A number of rather fragmentary observations of the infrared absorption spectrum of globular proteins, and of the infrared dichroism of protein crystals, have been held to provide some evidence for the presence of α-helices (for a summary of this evidence, see Doty and Geiduschek, 1953). This evidence was never very strong, and has now been shown to be virtually worthless by the discovery that those values of the absorption frequencies which had been held to be characteristic of the α-helix are also given by materials now known definitely not to possess this configuration. It would appear, in fact, that there may be no characteristic and diagnostic infrared frequencies for the α-fold, though such probably do exist for the β-fold (see Elliott and Malcolm, 1956c).

Measurements of optical rotation have recently been taken as an indication of the presence of α-helices in proteins and synthetic polypeptides. This development has resulted from the theoretical studies of Fitts and Kirkwood (1956 a, b) and of Moffitt and Yang (Moffitt, 1956a; Moffitt and Yang, 1956) and from the experimental investigations of Doty and his colleagues (Doty *et al.*, 1957) and of Elliott and his colleagues (Elliott *et al.*, 1956). In a recent paper Yang and Doty (1957) have found that the specific rotation and rotatory dispersion of a number of proteins are in accordance with the hypothesis that α-helices are present, and in the right-handed configuration, but by no means *all* the polypeptide chain in the molecule could have helical configuration. They quote, for example, 35–38% in insulin and 14–20% in ribonuclease—figures which may change considerably if the solvent is altered. In our view this type of evidence is suggestive but falls far short of being conclusive. It leads to a strong presumption that some sort of helical configuration is present, and of a single hand, but as far as we know does not discriminate between the α-helix and other helical configurations of a similar kind which have from time to time been proposed (e.g. the π-helix of Low and Baybutt, 1952; Low and Grenville-Wells, 1953). It may be remarked that there is an encouraging parallelism between the X-ray and optical results. For example, tropomyosin gives a strong 1.5 A reflection, and is one of the few proteins believed to contain a rather high percentage of α-helix (70–80%; Cohen and Szent-Györgyi, 1957); again, the X-ray results suggest that ribonuclease contains an unusually small amount of α-helix or other regularly folded structure (see p. 178), and similarly this protein is put well down the list on the basis of its optical rotation.

We conclude that, although it is plausible and attractive to suppose that globular proteins are made up in part of α-helices, the case is not yet proved. On the other hand it is undoubtedly useful as a working hypothesis, although it is virtually certain that if α-helices form the major part of globular proteins they are not all strictly parallel, or at least not in hemoglobin, myoglobin, or ribonuclease; otherwise the fact would have been apparent long ago. Whether certain globular proteins contain nonparallel α-helices—and we have seen that this is a plausible way of packing them (p. 165)—remains to be seen. Readers interested in the more controversial aspects of the subject may wish to refer to the papers of Bragg *et al.*, (1952) on hemoglobin; Carlisle and Scouloudi (1951) on ribonuclease; and Kendrew and P. J. Pauling (1956) and Kendrew and Parrish (1956) on myoglobin. Interesting models of insulin, based on mixtures of left-handed and right-handed α-helices, have been proposed by Linderstrøm-Lang and Schellman (1954), Lindley and Rollett (1955), and Low (1955). These are structurally plausible, but it is not yet known which is favored by the X-ray evidence, and to what extent.

It is generally assumed that since regular folds, including the α-helix, cannot accommodate proline residues (which, being imino-acid residues, have no spare hydrogen atom on the peptide bond to form a hydrogen bond), the polypeptide chains may kink or bend where the prolines occur. Another possible way of turning corners has been proposed by Lindley (1955). The fact is that although it is almost certain that the chains in globular proteins do turn corners—myoglobin, for example, consists of only a single polypeptide chain, so it must certainly be folded back on itself several times—there is no direct evidence how this occurs. The solution of the structure of one or two globular proteins will inject some life into these rather pale and wandering speculations. It may well be that the next review article on these topics in these volumes will be mainly concerned with such problems.

V. Viruses

Twenty years ago it was a matter for surprise that one could sometimes make crystals of viruses. The implications of the detailed diffraction patterns given by such virus crystals have not always been realized.

Crystals large enough to be seen have so far been obtained from comparatively few viruses, perhaps because the small amounts of material usually available have discouraged attempts at crystallization. This may be the reason why few crystals of animal viruses have been reported, although it seems unlikely that the larger and more irregular viruses, such as vaccinia, will ever be crystallized.

But even if only minute quantities are available a virus can be studied by the electron microscope (see the review by Williams, 1954). The remarkable fact has emerged that almost all *small* viruses have a fixed size and are, in shape, either rods or "spheres" (recent work has indicated that the "spheres" may be more nearly polyhedra; see, for example, Kaesberg, 1956). It is convenient to discuss the X-ray diffraction of virus particles under these two headings.

1. Rod-shaped Viruses

The most important rod-shaped virus, and the first to be examined by X-rays, is tobacco mosaic virus (TMV). This virus has been extensively studied by many different techniques and illustrates very well their various advantages and limitations. Here we shall naturally concentrate on the X-ray results, describing first of all the main features of the virus, then summarizing the X-ray studies, and finally commenting on the relation between the X-ray results and those obtained by other techniques.

a. General Features of TMV. The TMV particle is approximately cylindrical in shape.[5] Its length is 3000 A. and its mean diameter 150 A. It is made up of about 6% ribonucleic acid (RNA) and 94% protein. Chemical results strongly suggest that the protein component of the virus consists of rather over 2000 identical subunits, each of molecular weight about 18,000. Various strains of the virus are known; they generally differ slightly in the amino acid composition of their protein components. The RNA and the protein can be separated, and under favorable conditions these separated components can be recombined to give rodlike particles similar to the original virus. Moreover some of them appear to be infective (Fraenkel-Conrat and Williams, 1955). There is evidence, furthermore, that the separated RNA has some infectivity on its own (Gierer and Schramm, 1956). The protein component, known as "A" protein, usually has a molecular weight around 100,000 (and therefore must consist of an aggregate of several of the above-mentioned chemical subunits). It will polymerize at low pH to form rods similar to the intact virus, but of indefinite length (Schramm, 1947). *The virus protein is never infective in the absence of RNA.* Recently it has been found that RNA from sources other than TMV, and also various kinds of synthetic polynucleotides, can be co-aggregated with "A" protein to give (noninfectious) rods (Hart and Smith, 1956).

[5] The sources for most of the statements made in this section can be found in the volumes of *Advances in Virus Research* (K. M. Smith and M. A. Lauffer, eds., Academic Press, New York) and especially in the articles by C. A. Knight and R. C. Williams in Vol. 2, 1954.

True crystals of TMV, with sharp faces, are found within the cells of the host plant. They are too small to be examined by X-rays, but they have been studied in visible light by Wilkins *et al.* (1950) using a variety of techniques. These workers suggested that within the crystal the virus rods are arranged in rows with their ends in line, and that alternate rows of rods have slightly different orientations. This ingenious picture has recently been conformed by electron microscopy (Steere, 1957).

No one has so far succeeded in producing true crystals from *extracted* virus particles, but they easily form birefringent gels which are paracrystalline: that is to say, the virus rods are all parallel and in hexagonal array, but their ends are not lined up. These gels can swell and shrink, and the X-ray evidence strongly suggests that this behavior is a consequence, not of a change within the virus particle, but of changes in the distance between them.

b. X-ray Results: Basic Features. The most striking conclusion from the pioneer X-ray studies of Bernal and Fankuchen (1941) was that the virus is made up of subunits. This followed from the demonstration that the repeating unit in the direction of the virus axis was only 69 A. long, whereas the particle was known to be 3000 A. in length. The exact arrangement of these subunits was not clear at the time, but the essential feature was later suggested by Watson (1954), who pointed out that the X-ray pattern (in particular the region near the meridian where reflections are absent) could be most easily explained if the virus had a noninteger screw axis. The screw axis postulated had $(n + 1/3)$ subunits per turn, one complete turn occupying 23 A. It is difficult to determine the value of n from the diffraction data, and early suggestions that it might be 10 or 12 were based on a very insecure argument. Recent work, described below, make it likely that n is 16.

It was at one time thought that the virus might have a dyad axis of symmetry perpendicular to the screw axis, but this now seems very unlikely.

The postulate of a noninteger screw axis has been completely confirmed by the work of Dr. Rosalind Franklin and her co-workers (Franklin, 1955a, 1956a; Franklin and Klug, 1955). The clearest piece of evidence comes from a strain of virus known as U2, for which the parameters of the screw axis are slightly different, so that the structure does not *exactly* repeat after three turns. The subtle changes introduced into the diffraction pattern by this alteration are such as to be inexplicable on any other basis.

A number of other "strains" of TMV have been studied by X-rays (Franklin, 1956a). They all give extremely similar X-ray pictures, though minor differences can be detected. In particular the so-called cucumber virus 4 has a structure very close to that of TMV, though it has a slightly smaller mean diameter (146 A. instead of 152 A.). It is clear that all these "strains," including cucumber virus 4, are structurally related, but exactly how close this relationship may be biologically is another matter (Knight, 1955).

The X-ray pattern of reaggregated "A" protein (without RNA) is similar to that of TMV but less perfect, suggesting that the structure is basically the same but a little more irregular (Franklin, 1955b). The small differences between the patterns, probably due to the RNA, are discussed below. The change in the birefringence, from low positive for the intact virus to low negative for reaggregated protein, shows that the RNA makes a positive contribution to the birefringence of TMV. This result is compatible with earlier studies on the ultraviolet dichroism (Seeds and Wilkins, 1950), which established that the nitrogen bases of the RNA were arranged with their planes roughly parallel to the fiber axis, rather than perpendicular to it. It is interesting to note that when gels of reaggregated "A" protein are dried the layer line spacing shortens from 69 A to 62 A, whereas in the virus itself it does not change, presumably because the structure is constrained by the RNA (Franklin 1955b).

We cannot do more than barely mention the fact that certain globular proteins, immunologically and otherwise related to TMV, are found in infected plants (Rich *et al.*, 1955; Franklin and Commoner, 1955) and that these too are capable of aggregation into rodlike particles which give X-ray patterns somewhat resembling those from TMV.

c. X-ray results: the Internal Structure. A knowledge of the screw symmetry does not by itself tell us anything about the shape of the asymmetric unit, or the location of the RNA. Some information on these points has come from investigations using other methods of attack.

The first of these is the very careful work of Caspar (1955, 1956b) on the intensities of the equatorial reflections, leading to a direct deduction of their signs.

It can be shown that these reflections correspond to the cylindrical average of the electron density of the virus (at least if reflections of short spacing are excluded) and that the amplitudes of the reflections must be either positive or negative (i.e. the phase angle must be 0 or π). Caspar studied the first ten maxima of the intensity distribution, and showed that all sign combinations but two were very unlikely, in that they indicated a particle of too great a radius, and that of the two, one was distinctly better than the other.

His next approach was totally different, consisting in the application of the method of isomorphous replacement to a virus for the first time.

Lead was bound to TMV by adding to the mother liquor an amount of lead acetate corresponding to about 2500 lead atoms per virus particle (greater lead concentrations produced a curdy agglomerate). It could be deduced that the lead atoms were bound at two distinct radial distances from the virus axis, namely 25.3 A. and 84 A. Using this information it was possible to determine the signs of the reflections, and the result was

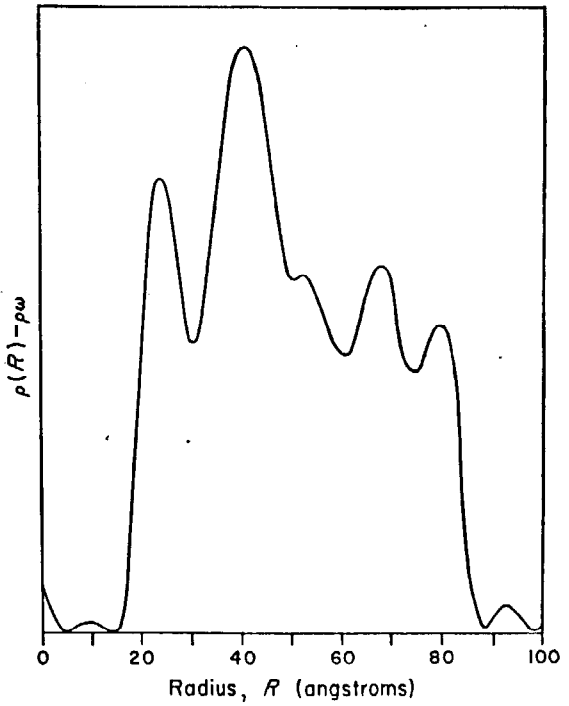

FIG. 21. Radial electron-density distribution in the tobacco mosaic virus particle, plotted as a function of distance from the axis of the particle. The density is the mean density in excess of that of water. Note the hole (filled with water) near the axis (R less than 20 A), and the large peak at the radius of 40 A. (Caspar, 1956b.)

identical with the preferred sign combination derived by the first method. Though the agreement between observed and calculated data was very good, some doubt might have been felt about Caspar's result because of the necessity of invoking *two* sites for the lead; however, recent work (see below) has confirmed his choice of signs.

The Fourier synthesis computed from the observed amplitudes, together with Caspar's chosen signs, shows the radial distribution of *average density* in the particle and is given in Fig. 21. Its most important features are the central minimum, representing a hole down the middle of the virus

(occupied by water), the peak at a radius of 24 A. together with the even larger peak at 40 A., and finally the fact that the radius of the particle appears to exceed the mean value of 75 A which had been deduced from earlier data.

Another important advance (Franklin, 1956b) has resulted from a second successful isomorphous replacement on the virus. Franklin studied a mercury substituted TMV, prepared by Fraenkel-Conrat, and containing one mercury per 20,000 molecular weight of protein. This proved to be isomorphous with unsubstituted TMV and a study of its

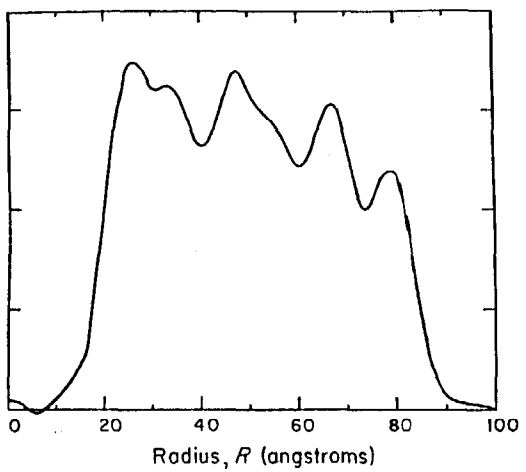

FIG. 22. Radial electron density distribution for repolymerized, RNA-free, "A" protein from TMV. Compare Fig. 21, which shows the corresponding function for intact TMV. The main difference is the absence here of a peak at 40 A radius. This, and other evidence, suggests that the RNA of the virus is located at a distance of 40 A from the axis of the virus particle. (Franklin, 1956b.)

diffraction pattern has confirmed Caspar's allocation of signs for the equatorial reflections, besides showing that the number of asymmetric units per turn is probably 16. It has also proved possible to allocate signs to the equatorial reflections of the aggregated "A" protein (free of RNA). The Fourier syntheses prepared using on the one hand the amplitudes of the equatorial reflections of the complete virus, and on the other those of the repolymerized "A" protein, show clearly that the RNA must be located at a radius of 40 A., since the only major difference between the two syntheses is that the large peak at 40 A. in the former is absent in the latter (see Figs. 21 and 22). Moreover the nature of the intensity differences in the first eight nonequatorial layer lines of the pattern all confirm that the RNA is located at a radius of about 40 A. Further work

is in progress which may reveal the general disposition of the RNA chain (or chains). Notice that the inner peak of Fig. 21, at a radius of 25 A., still appears when no RNA is present (Fig. 22) and must therefore be due to protein; but little so far is known about the disposition of the protein, in particular the arrangements of the polypeptide chains, though the diffraction data suggest that they may run perpendicular to the axis of the virus. The infrared dichroism of oriented TMV (Fraser, 1952) is compatible with this idea.

Franklin and Klug (1956) have reached some interesting conclusions about the external shape of the virus by making an entirely different approach to the data. When the virus particles are packed closely together (i.e. at a distance apart of 150 A.) the nature of the diffuse reflections in the pattern suggests that the virus surface is not smooth, but grooved or serrated; this produces helical disordering, as if one were packing together a set of screws. Moreover the intensity distribution on the third layer line suggested to them that there was some matter outside the *average* radius of 75 A. Their arguments, though not entirely compelling, are very suggestive and are moreover compatible with the conclusions reached by isomorphous replacement.

In summary, then, the X-ray studies of TMV have shown: (a) the virus is made up of identical (or at least very similar) protein subunits, related by a noninteger screw axis; (b) the surface of the virus is probably not smooth, but is grooved or serrated; (c) there is a hole of radius 20 A. down the center of the virus. The RNA is located near a radius of about 40 A.; some of the protein is at a smaller radius than this.

It is indeed remarkable that two successful isomorphous replacements should have been achieved in a virus at this relatively early stage in the application of the method to large molecules. Part of this success must be attributed to the high technical quality of the work of Caspar and Franklin.

The only rod-shaped virus unrelated to TMV which has been studied by X-rays is potato virus X. This gives much poorer X-ray patterns; and although they suggest that the structure is helical, better pictures will be needed before this can be established with certainty (Watson, unpublished data).

d. Correlation between X-ray and Other Results. We shall restrict ourselves to a comparison with electron microscopy and the chemical methods.

The electron microscope shows that the length of the virus is close to 3,000 A. This distance is too great to be resolved as a long-spacing X-ray reflection, but by using Bragg reflection of *visible light* Wilkins *et al.* (1950) have found that a spacing of this order does exist in intracellular virus

crystals. The *packing* diameter of TMV, measured on electron micrographs, is 150 A., but it is now realized that the diameter of *individual* virus particles in electron micrographs is a little greater than this. This agrees with the X-ray results which also show that the maximum diameter is a little greater than the mean diameter, and that the virus particles pack closely together by intermeshing.

By studying partially degraded virus in the electron microscope it can be seen that the RNA is near the axis of the particle (Hart, 1955; Schramm *et al.*, 1955). This does not agree with the X-ray results, according to which the RNA is at a radius of 40 A.: the discrepancy is probably due to the RNA collapsing toward the center when some of the supporting protein is removed preparatory to making electron micrographs.

On other electron micrographs pieces of "A" protein, shaped like disks, can be seen to have a hole in the middle (Hart, 1955; Fraenkel-Conrat and Williams, 1955). More recently Huxley (1957) has "stained" intact TMV with salts such as KCl, and has been able to demonstrate the hole down the center of the virus.

Although various claims have been made, no satisfactory demonstration of surface structure has yet been given by electron microscopy.

The chemical studies have shown that the protein of the virus is made up of small protein molecules whose molecular weight is about 18,000 (see the review (1956) by Anfinsen and Redfield). It has not yet been shown that all the subunits in a virus are identical, but they are certainly similar, for the amino acid sequences near the two ends of the single polypeptide chain show no signs of inhomogeneity. It has also been found that (with one possible exception) the various strains differ in their amino acid composition, sometimes quite strikingly; on the other hand, as might be expected, the X-ray results reveal only very minor differences between most strains, and these are in any case hard to interpret.

To summarize, the electron microscope has the advantage in studying large features. It is also a very valuable auxiliary tool since it needs such small amounts of material. The X-ray approach is unrivaled in studying structure at a somewhat higher resolution, and can also pick up features inside the intact virus. To detect differences at the amino acid level the chemical techniques are unequaled. The three methods, when properly used, give results which fit together into a coherent picture.

In particular we can combine data from all three to estimate the molecular weight. Accepting that there are 49 subunits per 69 A. of length, as indicated by the X-ray data, then in the total length of 3000 A. (measured in electron micrographs) there must be 2130 subunits, assuming no shrinkage in length when the virus dries in the electron microscope. The chemical methods suggest that the molecular weight of the protein subunit

is about 18,000. Allowing for 6% RNA these figures lead to a molecular weight of 40×10^6 for the whole virus. This figure should be compared with the value previously regarded as the best, namely 50×10^6, found by particle counts in electron micrographs (Williams *et al.*, 1951); actually earlier determinations by other methods had given figures nearer to 40×10^6. The agreement is only fair, but subsequent work may improve it.

2. Spherical Viruses

X-ray studies of spherical viruses are less advanced than those of TMV which we have just described, mainly because up to now no isomorphous replacement has been achieved. Merely noting that single crystal X-ray photographs of a spherical virus (Rothamsted strain of tobacco necrosis virus) were taken as early as 1945, by Crowfoot and Schmidt, we shall at once proceed to describe recent work on tomato bushy stunt virus and turnip yellow mosaic virus. For electron microscope studies see Williams (1954) and Kaesberg (1956).

a. Tomato Bushy Stunt Virus. This virus contains 17% of RNA by weight, the remainder being protein. Early X-ray studies (see Carlisle and Dornberger, 1948) showed that the unit cell was cubic in *shape* (a = 386 A.), but did not establish definitely that its *symmetry* was cubic. Caspar (1956a) has recently produced new evidence which suggests very strongly that the symmetry is indeed cubic, the space group being I23. There is only one molecule in the (primitive) unit cell; it follows that the virus particle itself must have cubic symmetry, its point group being 23, and therefore that it is made up of 12 identical subunits (see Table I). Following a careful study of the distribution of the strong reflections in reciprocal space Caspar has put forward very suggestive arguments that the point group may actually be of higher symmetry than the space group demands, namely 532 rather than merely 23: the 532 point group has 60 subunits, or a multiple thereof. Nothing has so far been discovered about the location of the RNA.

b. Turnip Yellow Virus. This material is of considerable interest because it was discovered (Markham, 1951) that the infective virus is accompanied in the plant by particles which, though otherwise similar to it, are noninfective and RNA-free. The infective virus contains about 40% RNA, the remainder being protein. The associated noninfective particle is 40% lighter, lacking as it does all RNA, yet its diameter is approximately the same (about 280 A.); its protein is similar immunologically to the protein component of the complete virus. The two particles form similar crystals, and will indeed form mixed crystals (Bernal and Carlisle, 1948). Furthermore, the low angle X-ray scattering of the RNA-free particle in solution, unlike that of the infective virus, is what one

would expect from a spherical shell rather than from a solid sphere (Schmidt *et al.*, 1954). Taking all this evidence together the conclusion is clear that by and large the protein is *outside* and the RNA *inside* the virus particle.

Some early electron micrographs of a few layers of virus (Cosslett and Markham, 1948) suggested that the lattice was of the diamond type, and the first X-ray studies were thought to confirm this. The problem has recently been taken up again by Klug, Finch, and Franklin, who kindly allowed us to see their manuscript prior to publication.

Their new data show clearly that the lattice has cubic symmetry, but are difficult to reconcile with the idea of a diamond lattice: the alternative would be a body-centered cubic lattice. A diamond lattice is a very open one, containing exactly half as much virus as would a body-centered cubic lattice in which the distance between neighboring virus particles was the same. It is thus possible to decide between the two alternatives simply by finding how much virus there is per unit volume of crystal. This has been done (in collaboration with Dr. Peter Walker) using a combination of ultraviolet absorption and interference microscopy; the result shows clearly that the more dense lattice is correct, at least as far as large crystals are concerned.

However a straightforward body-centered lattice cannot adequately explain the X-ray pattern except at low resolution. It is suggested that the *centers* of the virus particles fall on a simple body-centered cubic lattice, but that the particles may have either of two orientations which alternate in a regular manner; this means that the true unit cell is larger than the simple one. The same authors find that most of the strong intensities are located in positions in reciprocal space which correspond to the virus particle having the point group symmetry 532. Nevertheless one or two reflections are present which are quite incompatible with the virus having such a high symmetry, and Klug and his colleagues are driven to conclude that although the virus as a whole can only have 23 symmetry, some part of it may have the higher 532 symmetry.

This ingenious interpretation is too intricate to be completely accepted without further work, but there seems to be little doubt that turnip yellow virus has cubic symmetry of some sort. Results from the RNA-free protein component are eagerly awaited.

3. *General Principles of Virus Structure*

The fact that certain small viruses form crystals, and that these crystals in some cases give X-ray diffraction patterns extending to relatively small spacings (say 5 A.), shows quite clearly that such viruses can be loosely considered as "molecules" in the sense used by protein crystallographers; namely as entities in which the majority of the atoms are arranged in fixed (relative) positions. As we have indicated earlier, this does not carry the implication that the positions of *all* the atoms are fixed, nor that they are exactly the same in each virus; but it *does* imply that there is a very considerable similarity between the atomic arrangements of any

two sister virus particles. In the same way the existence of internal symmetry elements in the virus particle shows that one of its subunits must resemble any other subunit, though once again it is structural similarity rather than chemical identity which is proved by the X-rays.

The fact that small viruses are either rods or spheres (and not, for example, ellipsoids or plates) has suggested the hypothesis (Crick and Watson, 1956a) that they are all made of subunits, related by symmetry elements. This is a very natural idea to a crystallographer and had been proposed earlier in special cases (Hodgkin, 1949; Low, 1953), but it had not been sufficiently appreciated by virus workers themselves.

Reasons can be given why small viruses are made of subunits, but the arguments are speculative (see Crick and Watson, 1956b); however given that subunits do exist it is natural that we should find them to be related by symmetry elements. Such an arrangement means that every subunit has the same contact points with its neighbors—points at which it must be assumed that the same chemical groups are available in each of the identical subunits.

Apart from its *approximate* location in TMV and in turnip yellow virus, very little is known about the way the RNA is arranged in viruses or how it combines with the protein, except that the combination is unlikely to involve primary chemical bonds. It is nevertheless a very reasonable surmise that the RNA in the virus has the symmetry elements, or at least some of the symmetry elements, of the protein. In TMV this idea leads to the prediction (Crick and Watson, 1956a) that it is the *backbone* of the RNA which will follow this symmetry, not the sequence of the bases. This has been confirmed by the recent experiments of Hart and Smith (1956) who have shown that viruslike rods (of indefinite length) can be made by co-aggregating "A" protein from TMV with synthetic polyribotides having an RNA-like backbone, no matter what bases are attached to it. (That these polyribotides occupy the same sites in the "virus" particle as the native RNA does in the true virus is so far only an inference. It should be possible to prove it by X-ray methods.) It seems likely that the same prediction—that the backbone of the RNA possesses the same symmetry elements as the virus protein—will also be proved correct for the spherical viruses, but so far there is no evidence to support this.

Since symmetry elements can be discovered relatively easily by X-ray methods, and since they have been detected in all three plant viruses so far studied, it now becomes a worthwhile subject of inquiry whether any small virus under investigation has symmetry elements, and if so, what they are. However there is no law which says that a virus *must* have symmetry, and the hypothesis can only be evaluated by examining more and more types of virus; it remains to be seen whether the surmise of

Crick and Watson that *most* small "spherical" viruses have cubic symmetry will be confirmed or not.

Meanwhile the search for symmetry and the hope of isomorphous replacements (which would hardly be practicable if the virus did not contain identical subunits) are likely to stimulate an increasing amount of work in this field. Moreover X-ray investigation shows that at least certain viruses are simpler than their molecular weight might lead one to expect, and this should encourage further chemical studies, especially on the proteins of the spherical viruses.

It is not improbable (Crick and Watson, 1956b) that microsomal particles—the small compact particles in the cytoplasm which are perhaps the sites of protein synthesis—may also have cubic symmetry. They contain about the same amount of RNA as do the small spherical viruses, and have a similar (or perhaps slightly smaller) diameter, and they appear to be approximately spherical. It would not be surprising if the arrangement of the RNA were very similar in small viruses and in microsomal particles.

References

Ambrose, E. J., and Elliott, A. (1951). *Proc. Roy. Soc. London* **A205**, 47.

Ambrose, E. J., Bamford, C. H., Elliott, A., and Hanby, W. E. (1951). *Nature* **167**, 264.

Anfinsen, C. B., and Redfield, R. R. (1956). *Advances in Protein Chem.* **11**, 2.

Arndt, U. W., and Riley, D. P. (1955). *Phil. Trans. Roy. Soc. London* **A247**, 409.

Astbury, W. T., and Street, A. (1931). *Phil. Trans. Roy. Soc. London* **A230**, 75.

Astbury, W. T., and Woods, H. J. (1930). *Nature* **126**, 913.

Astbury, W. T., and Woods, H. J. (1933). *Phil. Trans. Roy. Soc. London* **A232**, 333.

Bamford, C. H., Brown, L., Elliott, A., Hanby, W. E., and Trotter, I. F. (1952). *Nature* **169**, 357.

Bamford, C. H., Brown, L., Elliott, A., Hanby, W. E., and Trotter, I. F., (1953). *Proc. Roy. Soc. London* **B141**, 49.

Bamford, C. H., Brown, L., Elliott, A., Hanby, W. E., and Trotter, I. F. (1954). *Nature* **173**, 27.

Bamford, C. H., Brown, L., Cant, E. M., Elliott, A., Hanby, W. E., and Malcolm, E. R. (1955). *Nature* **176**, 396.

Bamford, C. H., Elliott, A., and Hanby, W. E. (1956). "Synthetic Polypeptides." Academic Press, New York.

Bear, R. S. (1952). *Advances in Protein Chem.* **7**, 69.

Bear, R. S. (1955) in Fibrous Proteins and their Biological Significance. Symp. Soc. Exptl. Biol. IX, Cambridge Univ. Press.

Bear, R. S. (1956). *J. Biophys. Biochem. Cytol.* **2**, 363.

Bernal, J. D., and Carlisle, C. H. (1948). *Nature* **162**, 139.

Bernal, J. D., and Fankuchen, I. (1941). *J. Gen. Physiol.* **25**, 111, 147.

Bluhm, M. M., and Kendrew, J. C. (1956). *Biochim. et Biophys. Acta* **20**, 562.

Boedtker, H., and Doty, P. (1956). *J. Am. Chem. Soc.* **78**, 4267.

Bragg, W. L. (1939). "The Crystalline State," Vol. I. G. Bell & Sons Ltd. London.

Bragg, W. L., and Perutz, M. F. (1952a). *Acta Cryst.* **5**, 277.

Bragg, W. L., and Perutz, M. F. (1952b). *Acta Cryst.* **5**, 323.

Bragg, W. L., and Perutz, M. F. (1952c). *Proc. Roy. Soc. London* **A213**, 425.

Bragg, W. L., and Perutz, M. F. (1954). *Proc. Roy. Soc. London* **A225**, 315.

Bragg, W. L., Howells, E. R., and Perutz, M. F. (1952). *Acta Cryst.* **5**, 136.

Bragg, W. L., Howells, E. R., and Perutz, M. F. (1954). *Proc. Roy. Soc. London* **A222**, 33.

Brown, H., Sanger, F., and Kitai, R. (1955). *Biochem. J.* **60**, 556.

Brown, L., and Trotter, I. F. (1956). *Trans. Farday Soc.* **52**, 537.

Bunn, C. W. (1945). "Chemical Crystallography." Oxford Univ. Press, London and New York.

Carlisle, C. H., and Crowfoot, D. (1945). *Proc. Roy. Soc. London* **A184**, 64.

Carlisle, C. H., and Dornberger, K. (1948). *Acta Cryst.* **1**, 194.

Carlisle, C. H., and Scouloudi, H. (1951). *Proc. Roy. Soc. London* **A207**, 496.

Carlisle, C. H., Scouloudi, H., and Spier, M. (1953). *Proc. Roy. Soc. London* **B141**, 85.

Caspar, D. L. D. (1955). Ph.D. Thesis, Yale University.

Caspar, D. L. D. (1956a). *Nature* **177**, 475.

Caspar, D. L. D. (1956b). *Nature* **177**, 928.

Cochran, W., Crick, F. H. C., and Vand, V. (1952). *Acta Cryst.* **5**, 581.

Cohen, C., and Bear, R. S. (1953). *J. Am. Chem. Soc.* **75**, 2783.

Cohen, C., and Szent-Györgyi, A. (1957). *J. Am. Chem. Soc.* **79**, 248.

Corey, R. B., and Pauling, L. (1953). *Proc. Roy. Soc. London* **B141**, 10.

Coslett, V. E., and Markham, R. (1948). *Nature* **161**, 250.

Cowan, P. M., and McGavin, S. (1955a). *Comm. to 3rd Intern. Congr. Biochem.* **2**, 64.

Cowan, P. M., and McGavin, S. (1955b). *Nature* **176**, 501.

Cowan, P. M., McGavin, S., and North, A. C. T. (1955). *Nature* **176**, 1062.

Cowan, P. M., North, A. C. T., and Randall, J. T. (1953). *in* "The Nature and Structure of Collagen." Butterworth, London.

Cowan, P. M., North, A. C. T., and Randall, J. T. (1955) in Fibrous Proteins and their Biological Significance, Symp. Soc. Exp. Biol. IX, Camb. Univ. Press.

Crick, F. H. C. (1952). *Nature* **170**, 882.

Crick, F. H. C. (1953a). *Acta Cryst.* **6**, 689.

Crick, F. H. C. (1953b). Ph.D. Thesis, University of Cambridge.

Crick, F. H. C. (1954). *Sci. Progr.* **42**, 205.

Crick, F. H. C. (1956). *Acta Cryst.* **9**, 908.

Crick, F. H. C. (1957). *in* "Methods in Enzymology." (S. P. Colowick and N. O. Kaplan, eds.), Vol. IV. Academic Press, New York.

Crick, F. H. C., and Rich, A. (1955). *Nature* **176**, 780.

Crick, F. H. C., and Watson, J. D. (1956a). *Nature* **177**, 473.

Crick, F. H. C., and Watson, J. D. (1956b), Ciba Found. Symp. "The Nature of Viruses." Churchill, London.

Crowfoot, D., and Schmidt, G. M. S. (1945). *Nature* **155**, 504.

Donohue, J. (1952). *J. Chem. Phys.* **56**, 502.

Donohue, J. (1953). *Proc. Natl. Acad. Sci. U. S.* **39**, 470.

Doty, P., and Geiduschek, E. P. (1953). *in* "The Proteins." (H. Neurath and K. Bailey, eds.), Vol. I, Part A. Academic Press, New York.

Doty, P., and Lundberg, R. D. (1956). *J. Am. Chem. Soc.* **78**, 4810.

Doty, P., Bradbury, J. H., and Holtzer, A. M. (1956). *J. Am. Chem. Soc.* **78**, 947.
Doty, P., Wada, A., Yang, J. T., and Blout, E. R. (1957). *J. Polymer. Sci.* **23**, 851.
Elliott, A., and Malcolm, B. R. (1956a). *Trans. Farday Soc.* **52**, 528.
Elliott, A., and Malcolm, B. R. (1956b). *Nature* **178**, 912.
Elliott, A., and Malcolm, B. R. (1956c). *Biochim. et Biophys. Acta* **21**, 466.
Elliott, A., Hanby, W. E., and Malcolm, B. R. (1956). *Nature* **178**, 1170.
Fitts, D. D., and Kirkwood, J. G. (1956a). *Proc. Nat. Acad. Sci. U. S.* **42**, 33.
Fitts, D. D., and Kirkwood, J. G. (1956b). *J. Am. Chem. Soc.* **78**, 2650.
Fraenkel-Conrat, H., and Williams, R. C. (1955). *Proc. Natl. Acad. Sci. U. S.* **41**, 690.
Franklin, R. E. (1955a). *Nature* **175**, 379.
Franklin, R. E. (1955b). *Biochim. et Biophys. Acta* **18**, 313.
Franklin, R. E. (1956a). *Biochim. et Biophys. Acta* **19**, 203.
Franklin, R. E. (1956b). *Nature* **177**, 928.
Franklin, R. E., and Commoner, B. (1955). *Nature* **175**, 107.
Franklin, R. E., and Klug, A. (1955). *Acta Cryst.* **8**, 777.
Franklin, R. E., and Klug, A. (1956). *Biochim. et Biophys. Acta* **19**, 403.
Fraser, R. D. (1952). *Nature* **170**, 491.
Gierer, A., and Schramm, G. (1956). *Nature* **177**, 702.
Green, D. W., Ingram, V. M., and Perutz, M. F. (1954). *Proc. Roy. Soc.* **A225**, 287.
Green, D. W., North, A. C. T., and Aschaffenburg, R. (1956). *Biochim. et Biophys. Acta* **21**, 583.
Gustavson, K. H. (1955). *Nature* **175**, 70.
Harker, D. (1956). *in* "Biological and Medical Physics." (J. H. Lawrence and C. A. Tobias, eds.), Vol. IV. Academic Press, New York.
Hart, R. G. (1955). *Proc. Natl. Acad. Sci. U. S.* **41**, 261.
Hart, R. G., and Smith, J. D. (1956). *Nature* **178**, 739.
Hodgkin, D. C. (1949). *Cold Spring Harbour Symposia Quant. Biol.* **14**, 65.
Hodgkin, D. C., Kamper, M. J., Mackay, M., Pickworth, J., Trueblood, K. N., and White, J. G. (1956). *Nature* **178**, 64.
Huggins, M. L. (1952). *J. Am. Chem. Soc.* **74**, 3963.
Huxley, H. E. (1957). Electron Microscopy. Proc. Stockholm Conf. 1956, pp. 260, Almquist and Wiksell, Stockholm.
Huxley, H. E., and Kendrew, J. C. (1953). *Acta Cryst.* **6**, 76.
Ingram, V. M. (1955). *Biochem. J.* **59**, 653.
James, R. W. (1950). "X-ray crystallography," 4th ed. Methuen, London.
Kaesberg, P. (1956). *Science* **124**, 626.
Kendrew, J. C. (1950). *Proc. Roy. Soc. London* **A201**, 62.
Kendrew, J. C. (1954a). *Progr. in Biophys. and Biophys. Chem.* **4**, 244.
Kendrew, J. C. (1954b). *in* "The Proteins." (H. Neurath and K. Bailey, eds.), Vol. II, Part B. Academic Press, New York.
Kendrew, J. C., and Parrish, R. G. (1956). *Proc. Roy. Soc.* **A238**, 305.
Kendrew, J. C., and Pauling, P. J. (1956). *Proc. Roy. Soc.* **A237**, 255.
Kendrew, J. C., and Perutz, M. F. (1949). *in* "Haemoglobin" (Barcroft Memorial Volume). Butterworth, London.
Kendrew, J. C., and Perutz, M. F. (1957). *Ann. Rev. Biochem.* **27**, 327.
Kendrew, J. C., Parrish, R. G., Marrack, J. R., and Orlans, E. S. (1954). *Nature* **174**, 946.
King, M. V., Magdoff, B. S., Adelman, M. B., and Harker, D. (1956). *Acta Cryst.* **9**, 460.
Klug, A., Finch, J. T., and Franklin, R. E. (1957), in press.

Knight, C. A. (1955). *Virology* **1**, 261.
Krimm, S., and Schor, R. (1956). *J. Chem. Phys.* **24**, 922.
Kroner, T. D., Tabroff, W., and McGarr, J. J. (1955). *J. Am. Chem. Soc.* **77**, 3356.
Kupke, D. W., and Linderstrøm-Lang, K. (1954). *Biochim. et Biophys. Acta* **13**, 153.
Lang, A. R. (1956a). *Acta Cryst.* **9**, 436.
Lang, A. R. (1956b). *Acta Cryst.* **9**, 446.
Linderstrøm-Lang, K., and Schellman, J. A. (1954). *Biochim. et Biophys. Acta* **15**, 156.
Lindley, H. (1955). *Biochim. et Biophys. Acta* **18**, 194.
Lindley, H., and Rollett, J. S. (1955). *Biochim. et Biophys. Acta* **18**, 183.
Low, B. W. (1952). *J. Am. Chem. Soc.* **74**, 4830.
Low, B. W. (1953). *in* "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. I, Part A. Academic Press, New York.
Low, B. W. (1955). *Proc. Intern. Congr. Biochem. 3rd. Congr. Brussels*, p. 114.
Low, B. W., and Baybutt, R. B. (1952). *J. Am. Chem. Soc.* **74**, 5806.
Low, B. W., and Grenville-Wells, H. J. (1953). *Proc. Natl. Acad. Sci. U. S.* **39**, 785.
Magdoff, B. S., and Crick, F. H. C. (1955a). *Acta Cryst.* **8**, 461.
Magdoff, B. S., and Crick, F. H. C. (1955b). *Acta Cryst.* **8**, 468.
Magdoff, B. S., Crick, F. H. C., and Luzzati, V. (1956). *Acta Cryst.* **9**, 156.
Markham, R. (1951). *Discussions Faraday Soc.* **No. 11**, 221.
Marsh, R. E., Corey, R. B., and Pauling, L. (1955a). *Biochim. et Biophys. Acta* **16**, 1.
Marsh, R. E., Corey, R. B., and Pauling, L. (1955b). *Acta Cryst.* **8**, 62.
Marsh, R. E., Corey, R. B., and Pauling, L. (1955c). *Acta Cryst.* **8**, 710.
McMeekin, T. L., Rose, M. L., and Hipp, N. J. (1954). *J. Polymer. Sci.* **12**, 309.
Meggy, A. B., and Sikorski, J. (1956). *Nature* **177**, 326.
Meyer, K. H., and Go, Y. (1934). *Helv. Chim. Acta* **17**, 1488.
Moffitt, W. (1956a). *Proc. Natl. Acad. Sci. U. S.* **42**, 736.
Moffitt, W. (1956b). *J. Chem. Phys.* **25**, 467.
Moffitt, W., and Yang, J. T. (1956). *Proc. Natl. Acad. Sci. U. S.* **42**, 596.
Mustacchi, P. O. (1951). *Science* **113**, 405.
Palmer, K. T., Ballantyne, M., and Galvin, J. A. (1948). *J. Am. Chem. Soc.* **70**, 906.
Pauling, L., and Corey, R. B. (1951a). *Proc. Natl. Acad. Sci. U. S.* **37**, 241.
Pauling, L., and Corey, R. B. (1951b). *Proc. Natl. Acad. Sci. U. S.* **37**, 282.
Pauling, L., and Corey, R. B. (1951c). *Proc. Natl. Acad. Sci. U. S.* **37**, 729.
Pauling, L., and Corey, R. B. (1953a). *Nature* **171**, 59.
Pauling, L., and Corey, R. B. (1953b). *Proc. Natl. Acad. Sci. U. S.* **39**, 253.
Pauling, L., Corey, R. B., and Branson, H. R. (1951). *Proc. Natl. Acad. Sci. U. S.* **37**, 205.
Perutz, M. F. (1946). *Trans. Faraday Soc.* **B42**, 187.
Perutz, M. F. (1949). *Proc. Roy. Soc. London* **A195**, 474.
Perutz, M. F. (1951). *Nature* **167**, 1053.
Perutz, M. F. (1954). *Proc. Roy. Soc. London* **A225**, 264.
Perutz, M. F. (1956). *Acta Cryst.* **9**, 867.
Ramachandran, G. N. (1956). *Nature* **177**, 710.
Ramachandran, G. N., and Kartha, G. (1954). *Nature* **174**, 269.
Ramachandran, G. N., and Kartha, G. (1955). *Nature* **176**, 593.
Rich, A., and Crick, F. H. C. (1955). *Nature* **176**, 915.
Rich, A., Dunitz, J. D., and Newmark, P. (1955). *Nature* **175**, 1074.

Robertson, J. M. (1953). "Organic Crystals and Molecules." Cornell Univ. Press, Ithaca, N. Y.
Schellman, J. A. (1955). *Compt. rend. trav. lab. Carlsberg Ser. chim.* **29**, 230.
Schmitt, F. O., Hall, C. E., and Jakus, M. H. (1942). *J. Cellular Comp. Physiol.* **20**, 11.
Schmidt, P., Kaesberg, P., and Beeman, W. W. (1954). *Biochim. et Biophys. Acta* **14**, 1.
Schmitt, F. O., Gross, J., and Highberger, J. H. (1955). *Symposia Soc. Exptl. Biol.* **No. 9**, 148.
Schramm, G. (1947). *Z. Naturforsch.* **9b**, 779.
Schramm, G., Schumacher, G., and Zillig, W. (1955), *Nature* **175**, 549.
Schroeder, W. A., Kay, L. M., Le Gette, J., Hounen, L., and Green, F. C. (1954). *J. Am. Chem. Soc.* **76**, 3556.
Seeds, W. E., and Wilkins, M. H. F. (1950). *Discussions Faraday Soc.* **9**, 417.
Steere, R. L. (1957). *J. Biophys. Biochem. Cytol.* **3**, 45.
Trautman, R. (1956). *Abstr. Meeting, Am. Chem. Soc., Atlantic City.*
Tristram, G. R. (1953). "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. I, Part A. Academic Press, New York.
Warwicker, J. O. (1954). *Acta Cryst.* **7**, 565.
Watson, J. D. (1954). *Biochim. et Biophys. Acta* **13**, 10.
Wilkins, M. H. F., Stokes, A. R., Seeds, W. E., and Oster, G. (1950). *Nature* **166**, 127.
Williams, R. C. (1954). *Advances in Virus Research* **2**, 183.
Williams, R. C., Backus, R. C., and Steere, R. L. (1951). *J. Am. Chem. Soc.* **73**, 2062.
Yang, J. T., and Doty, P. (1957). *J. Am. Chem. Soc.* **79**, 761.